Dkt. 50950/JPW/EMW

Application for United States Letters Patent

To all whom it may concern:

Be it known that we, Wayne A. Hendrickson, Xuliang Jiang, Keith E. Langley, Rashid Syed and Yueh-Rong Ann Hsu

have invented certain new and useful improvements in

CONJUGATED LIGANDS FOR THE STIMULATION OF BLOOD CELL PROLIFERATION BY EFFECTING DIMERIZATION OF THE RECEPTOR FOR STEM CELL FACTOR

of which the following is a full, clear and exact description.

Dkt. 50950/JPW/EMW

CONJUGATED LIGANDS FOR THE STIMULATION OF BLOOD CELL PROLIFERATION BY EFFECTING DIMERIZATION OF THE RECEPTOR FOR STEM CELL FACTOR

Throughout this application, various references are referred to within parentheses. Disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains. Full bibliographic citation for these references may be found at the end of this application, preceding the claims.

Field of the Invention

This invention relates to stem cell factor (SCF) analogs, compositions containing such analogs, and related compositions. In another aspect, the present invention relates to nucleic acids encoding the present analogs or related nucleic acids, related host cells and vectors. In another aspect, the invention relates to computer programs and apparatuses for expressing the three dimensional structure of SCF and analogs thereof. In another aspect, the invention relates to methods for rationally designing SCF analogs and related compositions. In yet another aspect, the present invention relates to methods for treatment using the present SCF analogs.

BACKGROUND OF THE INVENTION

Stem cell factor (SCF) is an early-acting hematopoietic cytokine which elicits multiple biological effects. SCF is dimeric and occurs in soluble and membrane-bound forms. It transduces signals by ligand-mediated dimerization of its receptor, Kit. Kit is a receptor tyrosine kinase related to the receptors for platelet-derived growth factor (PDGF) and to those for vascular endothelial growth factor (VEGF), fibroblast growth factor (FGF), macrophage colony-stimulating factor (M-CSF) and Flt-3 ligand. The kinase portions of these receptors are closely related and their ligand-binding portions all comprise immunoglobulin-like (Ig) repeats, although these vary widely in sequence and also in number. Determined here is the crystal structure of selenomethionyl soluble human SCF at 2.2 Å resolution by multiwavelength anomalous diffraction (MAD) phasing. SCF has the characteristic helical cytokine topology, but the structure is unique apart from core portions. The SCF dimer has a symmetric 'head-to-head' association. Potential Kit-binding sites on the SCF dimer surface are located. A superposition of this dimer onto VEGF in its complex with the Flt-1 receptor places the binding sites on SCF in positions of topographical and electrostatic complementarity with the Kit counterparts of Flt-1. Similar models can be made for the complex of PDGF with its receptor and FGF-heparin.

INTRODUCTION

Stem cell factor (SCF) is an early-acting hematopoietic

cytokine that binds at the cell surface to its receptor, Kit, whereby it produces other biological effects in addition to those on hematopoiesis (see reviews by Galli et al., 1994; Lev et al., 1994; Besmer et al., 1997; Broudy, 1997). SCF, which is produced by various fibroblast-type cells including bone marrow stromal cells, has also been called Kit ligand (KL), mast-cell growth factor (MGF), and steel factor. The biochemistry and molecular biology that identified SCF and Kit as a ligand-receptor pair were preceded by an array of elegant animal biology studies that anticipated the underlying molecular mechanisms responsible for the genetics (Russell, 1979). Mice with mutations in the *Sl* locus (gene for SCF) or in the dominant-spotting *W* locus (*c-kit, the* gene for Kit) show complex phenotypes that include macrocytic anemia, sterility from a deficiency of germ cells, lack of coat pigmentation (white spotting of the skin from absences of pigment cells) and mast cell deficiency. Kit mutations in man are responsible for the autosomal dominant congenital pigmentation disorder, piebaldism. Consistent with these phenotypes, in the last 10 years, a host of *in vitro* and *in vivo* experiments have clearly established Kit-mediated roles for SCF in early stages of hematopoiesis, in gametogenesis, in melanocyte proliferation and function and in mast cell proliferation, maturation and activation; (Galli *et al.*, 1994, Lev et al., 1994, Besmer et al., 1997; Broudy, 1997). SCF has potential therapeutic applications in the treatment of anemias, boosting the mobilization of hematopoietic stem/ progenitor cells to the peripheral blood for harvest and transplantation, and in increasing

the efficiency of gene transduction for gene therapy (Galli *et al.*, 1994, McNiece and Briddell, 1995, Glaspy, 1996, Broudy, 1997).

Sub C1

SCF is expressed as membrane-associated forms of either 248 or 220 amino acid residues (Galli *et al.*, 1994, Lev et al., 1994, Besmer et al., 1997, Broudy, 1997) The two forms are a consequence of alternative mRNA splicing that includes or excludes exon 6. Exon 6 encodes a proteolytic cleavage site such that soluble SCF^{1-165} is released from the 248 amino-acid precursor. Residues 166-189 represent a tether to the membrane, residues 190-221 represent a hydrophobic transmembrane segment, and residues 222-248 represent a cytoplasmic domain. The 220 amino acid residue form lacks the cleavage site and tends to remain membrane-bound. Soluble SCF exists as a non-covalently associated dimer (Arakawa *et al.*, 1991). Each SCF monomer contains two intra-chain disulfide bridges, Cys4-Cys 89 and Cys43-Cys138 (Langley et al, 1992). The N-terminal 141 residues of SCF have been identified as a functional core, SCF^{1-141}, that includes the dimer interface and portions that bind and activate the receptor Kit (Langley *et al.*, 1994).

It has been proposed that SCF is a member of the helical cytokine structural superfamily characterized by a double-crossover four-helix bundle topology (Bazan, 1991). Three-dimensional structures are known for many of the family members and, from a comparison of the structures and sequences, the members have been classified into three subgroups (Sprang and Bazan, 1993):

short-chain, long-chain and interferon-like.

The superfamily is highly divergent. Among five short-chain helical cytokines of known structure, sequence identity levels rarely exceed 20% and fewer than half of the residues constitute (41%-48%) a common framework of the fold with r.m.s. deviations ranging from 1.7 Å to 2.9Å for the 61 C_α positions in common. Furthermore, many identical residues adopt different side chain conformations in the various structures. Nevertheless, sequence patterns do persist from the secondary structure and SCF has been proposed to be a short-chain helical cytokine (Bazan, 1991; Rozwarski *et al.*, 1994).

Most helical cytokines signal through members of the hematopoietic cytokine receptor superfamily, which are without intrinsic kinase activity (Heldin, 1995). SCF, in contrast, signals through a class III receptor tyrosine kinase (i.e. Kit). This class of kinases also includes the receptors for platelet-derived growth factor (PDGF), macrophage colony-stimulating factor (M-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), and Flt-3 ligand, and it is related to class V receptor tyrosine kinases (Flt-1, Flt-1/KDR and Flt-4) for vascular endothelial growth factors (VEGFs) (Fantl et al., 1993; Heldin, 1995; Rousset et al., 1995). The receptors in these classes have 'split' kinase domains intracellularly and multiple immunoglobulin(Ig)-like domains extracellularly.

The structures of PDGF (Oefner *et al.*, 1992), M-CSF (Pandit *et al.*, 1992), and VEGF (Muller *et al.*, 1997), have all been determined by X-ray crystallography, as has the complex of VEGF with domain 2 of its receptor, Flt-1 (Wiesmann *et al.*, 1997).

The ligands for the class III and class V receptors are all dimeric. As is the case for other ligands, SCF initiates signal transduction by dimerization of its receptor, Kit and the two juxtaposed receptors undergo tyrosine autophosphorylation (Heldin, 1995; Broudy, 1997), which initiates downstream intracellular signaling.

Here reported is the crystal structure of the core fragment of recombinant human stem cell factor, SCF^{1-141}, as determined at 2.2 Å resolution from multiwavelength anomalous diffraction (MAD) measurements. Incorporating data from mutagenesis and other structure-function studies, located were putative receptor-binding sites on the surface of the symmetric SCF dimer. From a comparison of these results with the structural and functional data for the related ligand-receptor systems, the complex of SCF with the receptor Kit is modeled and suggests a similar mode of association between other class III and class V receptors and their ligands.

Human SCF can be obtained and purified from a number of sources. SCF has been isolated from the rat and the mouse. Using the amino acid sequence of SCF protein isolated from the rat, the nucleic acid sequence encoding

the rat protein sequence was obtained from a rat cDNA library and then was cloned. The cloned nucleic acid encoding rat SCF was used to isolate, by hybridization, the nucleic acid molecule encoding human SCF from a human cDNA library. The development of recombinant DNA technology, see, for instance, U.S. Patent 4,810,643 (Souza) incorporated herein by reference, has enabled the production of commercial scale quantities of SCF in glycosylated form as a product of eukaryotic host cell expression, and of SCF in non-glycosylated form as a product of prokaryotic host cell expression.

SUMMARY OF THE INVENTION

The three dimensional structure of SCF has been determined herein to the atomic level. From this three-dimensional structure, one can now forecast with substantial certainty how changes in the composition of a SCF molecule may result in structural changes. These structural characteristics may be correlated with biological activity to design and produce SCF analogs.

This invention provides a computer based method for preparing a stem cell factor (SCF) analog comprising the steps of: (a) providing computer expression of the three dimensional structure of an SCF molecule using its crystal structure; (b) selecting from the computer expression of step (a) at least one site on the SCF molecule for alteration; (c) preparing an SCF molecule having an alteration at said at least one selected site; and (d) optionally, testing the SCF molecule for a desired characteristic.

This invention also provides an isolated SCF analog prepared according to the above-described computer based method for preparing a stem cell factor (SCF) analog comprising the steps of: (a) providing computer expression of the three dimensional structure of an SCF molecule using its crystal structure; (b) selecting from the computer expression of step (a) at least one site on the SCF molecule for alteration; (c) preparing a SCF molecule having an alteration at said at least one selected site; and (d) optionally, testing the SCF

molecule for a desired characteristic. In an embodiment the above-described SCF analog binds to SCF receptor, Kit. As used herein SCF receptor and "Kit" are used interchangeably to reflect the varied nomenclature used in the art.

This invention provides a composition comprising an isolated SCF analog prepared according to the above-described computer based method for preparing a stem cell factor (SCF) analog comprising the steps of: (a) providing computer expression of the three dimensional structure of an SCF molecule using its crystal structure; (b) selecting from the computer expression of step (a) at least one site on the SCF molecule for alteration; (c) preparing a SCF molecule having an alteration at said at least one selected site; and (d) optionally, testing the SCF molecule for a desired characteristic, effective to treat a subject and a pharmaceutically acceptable carrier.

This invention provides a method of treating a subject comprising administration of an isolated SCF analog prepared by the above-described computer based method for preparing a stem cell factor (SCF) analog comprising the steps of: (a) providing computer expression of the three dimensional structure of an SCF molecule using its crystal structure; (b) selecting from the computer expression of step (a) at least one site on the SCF molecule for alteration; (c) preparing a SCF molecule having an alteration at said at least one selected site; and (d) optionally, testing the SCF molecule for a

desired characteristic.

Sub C2

This invention provides a method for designing a compound (drug) capable of binding to the receptor of stem cell factor (SCF), Kit, comprising the steps of: a) determining a receptor binding site on the SCF based on the three dimensional structure of SCF or an SCF polypeptide capable of binding the receptor; and b) designing a compound comprising an entity that binds the SCF receptor. Accordingly, the designed compound is an SCF ligand analog, since a portion or part of the compound, "the entity", mimics the portion of SCF that binds to the SCF receptor, Kit. In step (a), and infra, the receptor binding site may be determined from atomic coordinates computed from X-ray diffraction data of a crystal comprising a polypeptide having an amino acid sequence portion of SCF capable of binding the receptor.

This invention provides a compound designed by the above-described method for designing a compound capable of binding to the receptor site of stem cell factor (SCF), Kit, comprising the steps of: a) determining a receptor binding site, on the SCF based on the atomic coordinates computed from X-ray diffraction data of a crystal comprising a polypeptide having an amino acid sequence portion of SCF capable of binding a ligand; and b) designing a **compound comprising an entity that binds the SCF receptor. As used herein, the entity, i,.e. the portion, of the designed compound fits the ligand binding site on the SCF receptor.**

This invention provides a method of treating a subject comprising administration of a compound designed by the above-described method for designing a compound capable of binding to the SCF receptor site.

This invention also provides a method of stimulating the production of hematopoietic calls in a subject comprising administering an isolated stem cell factor (SCF) analog or SCF ligand analogs to the subject.

Sub C3

This invention provides an isolated stem cell factor (SCF) molecule, which is an altered SCF, comprising any portion of amino acids 1-165 of a human SCF polypeptide, optionally comprising an N-terminal methionine before amino acid residue 1, wherein the polypeptide has an amino acid sequence portion of SCF capable of binding to the SCF receptor, Kit. Amino acid residue 1 of SCF is E, glutamic acid.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1C. Representative electron-density distributions in SCF. (Fig. 1A) MAD-phased experimental map calculated at 2.3 Å resolution. (Fig. 1B) The experimental map after four-fold averaging. (Fig. 1C) The current $2F_o$-F_c map superimposed with the model refined at 2.2 Å resolution. Each map is contoured at 1.0σ. Figures were drawn by the program O (Jones *et al.*, 1991).

Figures 2A-2B. Overall structure of an SCF dimer. (Fig. 2A) Ribbon diagram. (Fig. 2B) C_α stereodiagram of the AB dimer. Figures were drawn using the program SETOR (Evans, 1993).

SUB B'

~~**Figure 3.** Structure-based sequence alignment of SCF with other short-chain helical cytokines of human species. The dots denote gaps. M-CSF, IL-4, GM-CSF, IL-2 and IL-5 were aligned with SCF structure through structural superposition using TOSS (Hendrickson, 1979) and O (Jones *et al.*, 1991). C_α atoms were included if within 3.0 Å of their counterparts after superposition and at least three consecutive such residues are found in the fragment. The secondary structure~~

elements were assigned according to the output of the PROCHECK program (Laskowski *et al.*, 1993) except the helix assignment for residues 35-38, which was identified by inspection of the hydrogen-bond pattern. Secondary structures are shown in yellow with filled boxes referring to α-helices, half-filled boxes to 3_{10}-helices and arrows to β-strands. The solvent accessibility of the SCF dimer is indicated for each residue by an open circle if the fractional solvent accessibility is >0.4, a half-filled circle if it is 0.1-0.4, and a filled circle if it is <0.1. Residues at the SCF dimer interface are identified by stars, and the N-linked glycosylation sites by red Ys above the Asn residues.

Figures 4A-4B. Comparison of SCF dimer (shades of green) and M-CSF dimer (shades of brown). (Fig. 4A) View as in Figure 2. (Fig. 4B) View perpendicular to Fig. 4A, along the diad axis of M-CSF. Symmetry axes are shown as lines in Fig. 4A and dots in Fig. 4B. When one subunit of SCF dimer is superimposed onto a subunit of the M-CSF dimer, the other subunits are translated by 3.8 Å with a rotation of 4.7° to each other. Figures were generated using the program GRASP (Nicholls *et al.*, 1991).

SUB B2

Figure 5. Sequence alignments of SCF from human, mouse, rat and dog. (Anderson et al., 1990; Huang et al., 1990; Martin et al. 1990; Shull et al., 1992) The residues that are conserved in human and dog but different from rat and mouse are shadowed in yellow. Five regions of divergent sequence are identified (Roman numerals) Dots denote gaps, and dashes indicate residues identical to the human residues.

Figures 6A-6C. Ligand (worm structures)-receptor (yellow structures) models. (Fig. 6A) VEGF-Flt-1. (Fig. 6B) SCF-Kit. (Fig. 6C) PDGF-{DGF receptor. The used, without any modification, to approximate the receptor models. Receptor models are presented as yellow surfaces. The ligand models are presented as worm models. Background portions are colored light blue for one monomer and green for the other; receptor-interacting residues identified from site-directed mutagenesis experiments [VEGF (Muller et al.,1997), PDGF (Fenstermaker et al., 1993)] and other experimental data (SCF; see infra) arecolored magenta. Figures were drawn by the program GRASP (Nicholls *et al.*, 1991).

Figures 7A-7C. Electrostatic and carbohydrate surfaces of SCF and homology-modeled receptor Kit. (Fig. 7A) Electrostatic surface of SCF and worm of D2D3(Kit). (Fig. 7B) Electrostatic surface of Kit and worm of SCF. **(Fig. 8C)** Negative potential is colored red and positive potential, blue, with greatest saturations at -10 and +10kT, respectively. Carbohydrate moieties are represented by CPK models of a β-D-N-acetylglucose (green for SCF moieties and yellow for potential Kit moieties. Figures were drawn by the program GRASP (Nicholls *et al.*, 1991).

Figure 8. X-ray crystallographic coordinates of truncated stem cell factor molecule comprising amino acids 1-141 of a human SCF polypeptide.

Figure 9. Suggested renaming of the waters of the X-ray crystallographic coordinates set forth in Figure 8.

Figure 10. Design for a double-headed SCF ligand analog. (10A) General model (10B) Embodiment of the ligand head as an oligopeptide. The compound is the conjugation of a linker molecule with two ligand-head molecules. Each ligand head

is composed of up to three functional moieties, F_1, F_2 and F_3, which serve to mimic receptor-binding sites on the surface of SCF. Each ligand head also contains a conjugation moiety, F_L, endowed with chemical reactivity for conjugation with a reactive group at the end of the linker molecule. The capping moiety, F_C, at each end of the linker molecule is endowed with chemical reactivity for conjugation with the conjugation moiety from the ligand head. Double-headed molecules of this structure can have the property of binding to the SCF receptor, Kit, in such a way as to dimerize the receptor molecules and thereby lead to Kit activation in a manner analogous to the natural activity of SCF.

Ligand heads can be designed in at least four ways. (1) Ligand heads can be synthesized as oligopeptides wherein the functional moieties (F_1, F_2, F_3) are sequence elements from the SCF polypeptide; (2) The functional moieties (F_1, F_2, F_3) on such a ligand head can be selected by bacteriophage display for optimal receptor binding; (3) Chemical mimetics of the functional moieties and connecting segments in an active oligopeptide can be substituted for the respective moieties and segments; or (4)

An appropriate chemical framework (scaffold) of connecting segments can be designed to present functional moieties (F_1, F_2, F_3) which can be selected by combinatorial chemistry for optimal receptor binding from a library of chemical moieties complementary to receptor-binding sites on the surface of SCF.

When an oligopeptide embodiment of a linker head is designed in accord with option (1) it can have a sequence wherein F_1 corresponds to a segment from within the N-terminal region of SCF, residues 1-10; F_2 corresponds to a segment from within residues 79-95 (mainly located on the αC helix); F_3 is a segment from the C-terminal end of αD, near residue 127; F_L is a cysteine residue; and X_n, X_m, and X_p are connecting-peptide segments, composed from appropriate linker residues such as alanine, glycine, serine or proline, and wherein n=0-5, m=0-5 and p=3-8 residues, respectively.

Linkers can be designed from an organic polymer such as polyethylene glycol $H[OCH_2CH_2]_nOH$, where n=10-20 may suffice to separate the heads appropriately, wherein a reactive capping moiety, F_C, is appended at each end. The capping moiety may be a thiol reactive group, such as N-ethyl

maleimide, designed to bond covalently to the conjugation moiety, F_L, on the ligand head, wherein F_L may be cysteine residue or another thiol-containing group.

DETAILED DESCRIPTION OF THE INVENTION

The present determination of the three-dimensional structure to the atomic level is the most complete analysis to date, and provides important information to those wishing to design and prepare SCF analogs. For example, from the present three dimensional structural analysis, precise areas of hydrophobicity and hydrophilicity have been determined.

Relative hydrophobicity is important because it directly relates to the stability of the molecule. Generally, biological molecules, found in aqueous environments, are externally hydrophilic and internally hydrophobic; in accordance with the second law of thermodynamics provides, this is the lowest energy state and provides for stability. Although one could have speculated that SCF's internal core would be hydrophobic, and the outer areas would be hydrophilic, one would have had no way of knowing specific hydrophobic or hydrophilic areas. With the presently provided knowledge of areas of hydrophobicity/-philicity, one may forecast with substantial certainty which changes to the SCF molecule will affect the overall structure of the molecule.

As a general rule, one may use knowledge of the geography of the hydrophobic and hydrophilic regions to design analogs in which the overall SCF structure is not changed, but change does affect biological activity ("biological activity" being used here in its broadest sense to denote function). One may correlate biological

activity to structure. If the structure is not changed, and the mutation has no effect on biological activity, then the mutation has no biological function. If, however, the structure is not changed and the mutation does affect biological activity, then the residue (or atom) is essential to at least one biological function. Some of the present working examples were designed to provide no change in overall structure, yet have a change in biological function.

Sub C4

Based on the correlation of structure to biological activity, one aspect of the present invention relates to SCF analogs. These analogs are molecules which have more, fewer, different or modified amino acid residues from the SCF amino acid sequence. The modifications may be by addition, substitution, or deletion of one or more amino acid residues. The modification may include the addition or substitution of analogs of the amino acids themselves, such as peptidomimetics or amino acids with altered moieties such as altered side groups. The SCF used as a basis for comparison may be of human, animal or recombinant nucleic acid-technology origin (although the working examples disclosed herein are based on the recombinant production of the 141 amino acid species of human SCF, optionally having an extra N-terminal methionine residue). The analogs may possess functions different from natural human SCF molecule, or may exhibit the same functions, or varying degrees of the same functions. For example, the analogs may be designed to have a higher or lower biological activity, have a longer shelf-life or a decrease in stability, be easier to

Sub C4 cont'd

formulate, or more difficult to combine with other ingredients. The analogs may bind receptor but elicit no biological activity and may therefore be useful as an antagonist against SCF effect (as, for example, in the overproduction of SCF). From time to time herein the present analogs are referred to as proteins or peptides for convenience, but contemplated herein are other types of molecules, such as peptidomimetics or chemically modified peptides.

In embodiment, the present invention relates to related compositions containing a SCF analog as an active ingredient. The term, "related composition," as used herein, is meant to denote a composition which may be obtained once the identity of the SCF analog is ascertained (such as a SCF analog labeled with a detectable label or pharmaceutical composition). Also considered a related composition are chemically modified versions of the SCF analog, such as those having attached at least one polyethylene glycol molecule.

For example, one may prepare a SCF analog to which a detectable label is attached, such as a fluorescent, chemiluminescent or radioactive molecule.

Another example is a pharmaceutical composition which may be formulated by known techniques using known materials, see, e.g., Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, Pennsylvania 18042) pages 1435-1712, which are herein incorporated by reference. Generally, the formulation will depend on a

variety of factors such as administration, stability, production concerns and other factors. The SCF analog may be administered by injection or by pulmonary administration via inhalation. Enteric dosage forms may also be available for the present SCF analog compositions, and therefore oral administration may be effective. SCF analogs may be inserted into liposomes or other microcarriers for delivery, and may be formulated in gels or other compositions for sustained release. Although preferred compositions will vary depending on the use to which the composition will be put, generally, for SCF analogs having at least one of the biological activities of natural SCF, preferred pharmaceutical compositions are those prepared for subcutaneous injection or for pulmonary administration via inhalation, although the particular formulations for each type of administration will depend on the characteristics of the analog.

Another example of related composition is a receptor for the present analog. As used herein, the term "receptor" indicates a moiety which selectively binds to the present analog molecule. For example, antibodies, or fragments thereof, or "recombinant antibodies" (see Huse et al., Science 246:1275 (1989)) may be used as receptors. Selective binding does not mean only specific binding (although binding-specific receptors are encompassed herein), but rather that the binding is not a random event. Receptors may be on the cell surface or intra- or extra-cellular, and may act to effectuate, inhibit or localize the biological activity of the present analogs.

Receptor binding may also be a triggering mechanism for a cascade of activity indirectly related to the analog itself. Also contemplated herein are nucleic acids, vectors containing such nucleic acids and host cells containing such nucleic acids which encode such SCF analogs.

Another example of a related composition is a SCF analog with a chemical moiety attached. Generally, chemical modification may alter biological activity or antigenicity of a protein, or may alter other characteristics, and these factors will be taken into account by a skilled practitioner. As noted above, one example of such chemical moiety is polyethylene glycol. Modification may include the addition of one or more hydrophilic or hydrophobic polymer molecules, fatty acid molecules, or polysaccharide molecules. Examples of chemical modifiers include polyethylene glycol, alklpolyethylene glycols, DI-poly(amino acids), polyvinylpyrrolidone, polyvinyl alcohol, pyran copolymer, acetic acid/acylation, proprionic acid, palmitic acid, lecithin, stearic acid, dextran, carboxymethyl cellulose, pullulan, or agarose. See, Francis, *Focus on Growth Factors* 3: 4-10 (May 1992)(published by Mediscript, Mountview Court, Friern Barnet Lane, London N20 OLD, UK). Also, chemical modification may include an additional protein or portion thereof, use of a cytotoxic agent, or an antibody.

In another embodiment, the present invention relates to nucleic acids encoding such analogs. The nucleic acids

may be DNAs or RNAs or derivatives thereof, and will typically be cloned and expressed on a vector, such as a phage or plasmid containing appropriate regulatory sequences. The nucleic acids may be labeled (such as using a radioactive, chemiluminescent, or fluorescent label) for diagnostic or prognostic purposes, for example. The nucleic acid sequence may be optimized for expression, such as including codons preferred for bacterial expression. The nucleic acid and its complementary strand, and modifications thereof which do not prevent encoding of the desired analog are here contemplated.

In another embodiment, the present invention relates to host cells containing the above nucleic acids encoding the present analogs. Host cells may be eukaryotic or prokaryotic, and expression systems may include extra steps relating to the attachment (or prevention) of sugar groups (glycosylation), proper folding of the molecule, the addition or deletion of leader sequences or other factors incident to recombinant expression.

In further embodiment the present invention relates to antisense nucleic acids which act to prevent or modify the type or amount of expression of such nucleic acid sequences. These may be prepared by known methods.

In another embodiment of the present invention, the nucleic acids encoding a present analog may be used for gene therapy purposes, for example, by placing a vector containing the analog-encoding sequence into a recipient

so the nucleic acid itself is expressed inside the recipient who is in need of the analog composition. The vector may first be placed in a carrier, such as a cell, and then the carrier placed into the recipient. Such expression may be localized or systemic. Other carriers include non-naturally occurring carriers, such as liposomes or other microcarriers or particles, which may act to mediate gene transfer into a recipient.

The present invention also provides for computer programs for the expression (such as visual display) of the SCF or analog three dimensional structure, and further, a computer program which expresses the identity of each constituent of an SCF molecule and the precise location within the overall structure of that constituent, down to the atomic level. Set forth below is one example of such program. There are many currently available computer programs for the expression of the three dimensional structure of a molecule. Generally, these programs provide for inputting of the coordinates for the three dimensional structure of a molecule (i.e., for example, a numerical assignment for each atom of an SCF molecule along an x, y, and z axis), means to express (such as visually display) such coordinates, means to alter such coordinates and means to express an image of a molecule having such altered coordinates. One may program crystallographic information, i.e., the coordinates of the location of the atoms of an SCF molecule in three dimensional space, wherein such coordinates have been obtained from crystallographic analysis of said SCF molecule, into such programs to generate a computer

program for the expression (such as visual display) of the SCF three dimensional structure. Also provided, therefore, is a computer program for the expression of SCF analog three dimensional structure. Preferred is the computer program Insight II, version 4, available from Biosym, San Diego, California, with the coordinates as set forth in Figure 8 input. Preferred expression means is on a Silicon Graphics 320 VGX computer, with Crystal Eyes glasses (also available from Silicon Graphics), which allows one to view the SCF molecule or its analog stereoscopically. The above-listed computer programs are only examples, and the use of such programs in the claimed methods is not limited thereto, as one of skill may use any other computer program that provides the desired three dimensional expression. Alternatively, the present SCF crystallographic coordinates and diffraction data are also deposited in the Protein Data Bank, Chemistry Department, Rutgers University, New Jersey, USA [formerly at Brookhaven National Laboratory, Upton, NY 11972]. One may use these data in preparing a different computer program for expression of the three dimensional structure of a SCF molecule or analog thereof. Therefore, another aspect of the present invention is a computer program for the expression of the three dimensional structure of a SCF molecule. Also provided is said computer program for visual display of the three dimensional structure of an SCF molecule; and further, said program having means for altering such visual display. Apparatus useful for expression of such computer program, particularly for the visual display of the computer image of said three dimensional structure of

an SCF molecule or analog thereof is also therefore here provided, as well as means for preparing said computer program and apparatus.

The computer program is useful for preparation of SCF analogs because one may select specific sites on the SCF molecule for alteration and readily ascertain the effect the alteration will have on the overall structure of the SCF molecule. Selection of said site for alteration will depend on the desired biological characteristic of the SCF analog. If one were to randomly change said SCF molecule there would be substitutions, additions or deletions, and even more analogs having multiple changes. By viewing the three dimensional structure wherein said structure is correlated with the composition of the molecule, the selection for sites for alteration is no longer a random event, but sites for alteration may be determined rationally.

Identity of the three dimensional structure of SCF, including the placement of each constituent down to the atomic level has now yielded information regarding which moieties are necessary to maintain the overall structure of the SCF molecule. One may therefore select whether to maintain the overall structure of the SCF molecule when preparing an SCF analog of the present invention, or whether (and how) to change the overall structure of the SCF molecule when preparing a SCF analog of the present invention. Optionally, once one has prepared such analog, one may test such analog for a desired characteristic.

One may, for example, seek to maintain the overall structure possessed by a non-altered natural or recombinant SCF molecule. The overall structure is presented in Figures 2A-2B, and is described in more detail below. Maintenance of the overall structure may ensure receptor binding, a necessary characteristic for an analog possessing the biologic capabilities of natural SCF (if no receptor binding, signal transduction does not result from the presence of the analog). It is contemplated that one class of SCF analogs will possess the three dimensional core structure of a natural or recombinant (non-altered) SCF molecule, yet possess different characteristics, such as an increased ability to selectively stimulate neutrophils. Another class of SCF analogs are those with a different overall structure which diminishes the ability of an SCF analog molecule to bind to a SCF receptor, Kit, and possesses a diminished ability to selectively stimulate hematopoiesis, for example, as compared to non-altered natural or recombinant SCF.

For example, it is now known which moieties within the internal regions of the SCF molecule are hydrophobic, and, correspondingly, which moieties on the external portion of the SCF molecule are hydrophilic. Without knowledge of the overall three dimensional structure, preferably to the atomic level as provided herein, one could not forecast which alterations within this hydrophobic internal area would result in a change in the overall structural conformation of the molecule. An

overall structural change could result in a functional change, such as lack of receptor binding, for example, and therefore, diminishment of biological activity as found in non-altered SCF. Another class of SCF analogs is therefore SCF analogs which possess the same hydrophobicity as (non-altered) natural or recombinant SCF. More particularly, another class of SCF analogs possesses the same hydrophobic moieties within the four helical bundle of its internal core as those hydrophobic moieties possessed by (non-altered) natural or recombinant SCF yet have a composition different from said non-altered natural or recombinant SCF.

Another example relates to external loops which are structures which connect the internal core (helices) of the SCF molecule. From the three dimensional structure -- including information regarding the spatial location of the amino acid residues -- one may forecast that certain changes in certain loops will not result in overall conformational changes.

Therefore, another class of SCF analogs provided herein is that having an altered external loop but possessing the same overall structure as (non-altered) natural or recombinant SCF. More particularly, another class of SCF analogs provided herein are those having an altered external loop, said loop being selected from the loops discussed infra. More particularly, said loops, are altered to increase the half life of the molecule by stabilizing said loops. Such stabilization may be by connecting all or a portion of said loop(s) to a portion

of an alpha helical bundle found in the core of a SCF (or analog) molecule. Such connection may be via beta sheet, salt bridge, disulfide bonds, hydrophobic interaction or other connecting means available to those skilled in the art, wherein such connecting means serves to stabilize said external loop or loops.

Additionally, such external loops may be the site(s) for chemical modification because in (non-altered) natural or recombinant SCF such loops are relatively flexible and tend not to interfere with receptor binding. Thus, there would be additional room for a chemical moiety to be directly attached (or indirectly attached via another chemical moiety which serves as a chemical connecting means). The chemical moiety may be selected from a variety of moieties available for modification of one or more function of an SCF molecule. For example, an external loop may provide sites for the addition of one or more polymer which serves to increase serum half-life, such as a polyethylene glycol molecule. Such polyethylene glycol molecule(s) may be added wherein said loop is altered to include additional lysines which have reactive side groups to which polyethylene glycol moieties are capable of attaching. Other classes of chemical moieties may also be attached to one or more external loops, including but not limited to other biologically active molecules, such as receptors, other therapeutic proteins (such as other hematopoietic factors which would engender a hybrid molecule), or cytotoxic agents (such as diphtheria toxin). This list is of course not complete; one skilled in the art possessed of

the desired chemical moiety will have the means to effect attachment of said desired moiety to the desired external loop. Therefore, another class of the present SCF analogs includes those with at least one alteration in an external loop wherein said alteration provides for the addition of a chemical moiety such as at least one polyethylene glycol molecule.

Deletions, such as deletions of sites recognized by proteins for degradation of the molecule, may also be effectual in the external loops. This provides alternative means for increasing half-life of a molecule otherwise having the SCF receptor binding and signal transduction capabilities (e.g., the ability to selectively stimulate hematopoiesis). Therefore, another class of the present SCF analogs includes those with at least one alteration in an external loop wherein said alteration decreases the turnover of said analog by proteases. One may prepare an abbreviated SCF molecule by deleting a portion of the amino acid residues found in any of the the external loops (discussed infra), said abbreviated SCF molecule may have additional advantages in preparation or in biological function.

Another example relates to the relative charges between amino acid residues which are in proximity to each other. As noted above, the SCF molecule contains a relatively tightly packed four helical bundle. Some of the faces on the helices face other helices. At the point (such as a residue) where a helix faces another helix, the two amino acid moieties which face each other may have the same

charge, and thus tend to repel each other, which lends instability to the overall molecule. This may be eliminated by changing the charge (to an opposite charge or a neutral charge) of one or both of the amino acid moieties so that there is no repelling. Therefore, another class of SCF analogs includes those SCF analogs having been altered to modify instability due to surface interactions, such as electron charge location.

The present invention provides methods for designing SCF analogs and related compositions and the products of those methods. The end products of the methods may be the SCF analogs as defined above or related compositions. For instance, the examples disclosed herein demonstrate (a) the effects of changes in the constituents (i.e., chemical moieties) of the SCF molecule on the SCF structure and (b) the effects of changes in structure on biological function.

Accordingly, therefore, the present invention provides a computer based method for preparing a stem cell factor (SCF) analog comprising the steps of: (a) providing computer expression of the three dimensional structure of of an SCF molecule using its crystal structure; (b) selecting from the computer expression of step (a) at least one site on the SCF molecule for alteration; (c) preparing an SCF molecule having an alteration at said one said selected site; and (d) optionally, testing the SCF molecule for a desired characteristic. The SCF molecule of step (a) may be naturally occurring wild type SCF or any portion or fragment thereof which is capable

of binding to SCF receptor.

Sub C5

In an embodiment of the above-described method the computer expression allows for display of the amino acids of the SCF molecule. In another embodiment of the method the computer expression allows for display of each atom of the SCF molecule. In a further embodiment of the method the SCF molecule is a native or a selenomethionyl SCF. In another embodiment of the method the site on the SCF molecule for alteration is a receptor binding site on the surface of the SCF molecule. In a further embodiment of the method the receptor binding site comprises amino acid residues 79-85. The SCF molecule may be a recombinant human SCF or a wild type naturally occurring human SCF. SCF wild type and recombinant may also be of other sources such as but not limited to rat or mouse. In an embodiment of the above-described method, the atomic coordinates of the crystal structure are set forth in Figure 8. In another embodiment the SCF analog comprises a polypeptide having an amino acid sequence portion of SCF capable of binding a receptor and having the overall three-dimensional conformation as shown in Figures 2A-2B, wherein the three-dimensional conformation is: a) anti-parallel, double-cross over 4-alpha helical bundle with a left hand twist; and b) overall dimensions of approximately 85 Å x 30 Å x 20 Å. In an embodiment the SCF analog comprises electron density distributions as set forth in Figures 1A, 1B, and 1C. In a further embodiment the SCF molecule is a native SCF or a selenomethionyl SCF.

In an embodiment the site on the SCF molecule for

alteration is a receptor binding site on the surface of the SCF molecule or a non-receptor site of the SCF.

Alteration of a non-receptor binding site will result in a designed SCF analog that binds to the SCF receptor but is less active such that such an analog may be used for blocking activity of the SCF.

Sub C6

In another embodiment the receptor binding site comprises approximately amino acid residues 79-95.

This invention provides an isolated SCF analog prepared according to the above-described method. In an embodiment the isolated SCF analog which binds to SCF receptor, Kit. In another embodiment the isolated SCF analog has an alteration in at least one atom of the atomic coordinates of the crystal structure set forth in Figure 8. In a further embodiment the SCF analog comprises a polypeptide having an amino acid sequence portion of SCF capable of binding a receptor and having the overall three-dimensional conformation as shown in Figures 2A-2B, wherein the three-dimensional conformation is: a) anti-parallel, double-cross over 4-alpha helical bundle with a left hand twist; and b) overall dimensions of approximately 85 Å x 30 Å x 20 Å. In an embodiment the SCF analog comprises electron density distributions altered from those set forth in Figures 1A, 1B, and 1C.

Sub C7

This invention provides a composition comprising an isolated SCF analog prepared according to the above-described method effective to treat a subject and a

Sub C7 Cont'd

pharmaceutically acceptable carrier. In an embodiment of the composition, the isolated SCF analog has an alteration in at least one atom of the atomic coordinates of the crystal structure set forth in Figure 8. In another embodiment the isolated SCF analog comprises a polypeptide having an amino acid sequence portion of SCF capable of binding a receptor and having the overall three-dimensional conformation as shown in Figures 2A-2B, or an alteration thereof, wherein the three-dimensional conformation is: a) anti-parallel, double-cross over 4-alpha helical bundle with a left hand twist; and b) overall dimensions of approximately 85 Å x 30 Å x 20 Å. In a further embodiment the isolated SCF analog comprises electron density distributions as set forth in Figures 1A, 1B, and 1C. In an embodiment the isolated SCF analog comprises a native SCF1-165, a recombinant seleno-methionyl SCF1-141, or a recombinant selenomethionyl SCF1-165.

Any of the aforementioned SCF analogs may optionally have before the first N-terminal amino acid residue a methionine at position"-1".

Sub C8

In an embodiment of the composition the site on the isolated SCF molecule for alteration is a receptor binding site on the surface of the SCF molecule. In a further embodiment the receptor binding site comprises approximately amino acid residues 79-95.

Sub C9

This invention provides a method of treating a subject having a disorder requiring SCF comprising administration

Sub C9 Cont'd

of a composition comprising an isolated SCF analog prepared by the method of preparing a SCF analog or a compound designed by the method of designing a compound capable of binding to the SCF receptor as described infra. In an embodiment the subject has a blood disorder. In another embodiment the disorder which the subject has is anemia, myeloproliferative disorder, neoplasia, nerve damage, infertility, intestinal damage, a pigmentation disorder, or immunodeficiency. In an embodiment the administration of the isolated SCF analog is for ex vivo or in vivro production of peripheral blood progenitors, ex vivo or in vitro stem cell expansion, ex vivo or in vitro growth of epithelial cells, ex vivo or in vitro growth of stromal cells, ex vivo or in vitro dendritic cell stimulation, and in vivo cell mobilization. In an embodiment the isolated SCF analog is administered orally or by any other routes described infra. In an embodiment the isolated SCF analog has an alteration in at least one atom of the atomic coordinates of the crystal structure set forth in Figure 8. In a further embodiment the isolated SCF analog comprises a native SCF1-165 or a recombinant selenomethionyl SCF1-141. In another embodiment the site on the isolated SCF molecule for alteration is a receptor binding site on the surface of the SCF molecule. In a further embodiment the receptor binding site comprises approximately amino acid residues 79-95. In an embodiment the isolated SCF analog comprises a native or recombinant SCF1-165 or a recombinant selenomethionyl SCF1-141. As used herein throughout SCF receptor is Kit.

This invention provides a method for designing a compound capable of binding to the stem cell factor (SCF) receptor site of comprising the steps of: a) determining a binding site for the SCF receptor on the SCF based on the three-dimensional structure of SCF or an SCF polypeptide or portion/fragment thereof, atomic coordinates computed from X-ray diffraction data of a crystal comprising a polypeptide having an amino acid sequence portion of SCF capable of binding the receptor; and b) designing a compound comprising an entity that binds the SCF receptor. The designed compound mimics, i.e. is a copy or simulation of the overall portion of SCF that binds to SCF receptor, Kit.

In an embodiment the design of the compound of step (b) is determined by shape complementarity or by estimated interaction energy. In another embodiment the designed compound fits an SCF receptor binding site on SCF receptor as shown in Figure 6. In a further embodiment the designed compound fits an SCF receptor binding site on SCF receptor as shown in Figures 7A or 7B. In an embodiment the designed compound has an alteration in at least one atom of the atomic coordinates of the crystal structure set forth in Figure 8. In yet another embodiment the designed compound is a double-headed SCF ligand analog having the structure set forth in Figure 10A. In a still further embodiment each ligand head of the double-headed SCF ligand analog is an oligopeptide having the structure set forth in Figure 10B. The designed compound comprises two conjugated ligands having a linker between the two ligands.

Sub C10

In an embodiment, the ooligopeptide comprises a sequence, wherein functional moiety F_1 corresponds to a segment of amino acid residues from within N-terminal residues 1-10 of SCF, functional moiety F_2 corresponds to a segment of amino acid residues from within residues 79-95 of SCF, and functional moiety F_3 corresponds to a segment of amino acid residues located within three amino acid residues of amino acid residue 127, wherein F_1, F_2, and F_3 are connected by connecting peptide segements X_n, X_m, and X_p, respectively, wherein n=0-5, m=0-5 and p=3-8 amino acid residues, respectively, and the conjugation moiety F_L is a cysteine residue.

A functional moiety is defined as en entity that has a particular binding property, i.e. it mimics receptor-binding sites on the surface of SCF, i.e. the ligand portion of SCF.

Sub C11

The amino acid residues located within 3 amino acid residues of amino acid residue 127 may be located within 3 residues in either direction of residue 127. In further embodiments the amino acid residues may be from 4 to 10 amino acid residues in either direction of amino acid residue 127.

In another embodiment of the above-described method the functional moieties F_1, F_2, and F_3 on the ligand heads have been selected by bacterial phage display for optimal receptor binding. In an embodiment the functional moieties and connecting peptide segments of an active oligopeptide ligand head are replaced by chemical

mimetics. In another embodiment an appropriate chemical scaffold of connecting segments has been designed to comprise (present) functional moieties F_1, F_2 and F_3 which have been selected by combinatorial chemistry for optimal receptor binding from a library of chemical moieties complementary to receptor-binding sites on the surface of SCF. In an embodiment the linker comprises an organic polymer having two ends capped at each end by a reactive capping moiety, F_c, which react covalently with the conjugation moiety, F_L, on the ligand head. In an embodiment the organic polymer is polyethyleneglycol (PEG) comprising the structure $H[OCH_2CH_2]_nOH$, wherein n is 10-20. In an embodiment the capping moiety, F_c, is a thiol-reactive group such as N-ethyl maleimide. In an embodiment the conjugating moiety, F_L, is a thiol containing group such as cysteine.

This invention provides a compound designed by the method of claim 32.

A composition comprising the compound designed by the above described method and a pharmaceutically acceptable carrier. In an embodiment the compound comprises an isolated SCF analog, whose alteration site is a receptor-binding site on the surface of the altered SCF molecule. In another embodiment the composition comprises a double-headed receptor SCF ligand analog having the structure set forth in Figure 10A. In an embodiment each ligand head of the double-headed SCF ligand analog is an oligopeptide having the structure set forth in Figure 10B.

Sub C12

In another embodiment the ooligopeptide comprises a sequence, wherein functional moiety F_1 corresponds to a segment of amino acid residues from within N-terminal residues 1-10 of SCF, functional moiety F_2 corresponds to a segment of amino acid residues from within residues 79-95 of SCF, and functional moiety F_3 corresponds to a segment of amino acid residues located within three amino acid residues of amino acid residue 127, wherein F_1, F_2, and F_3 are connected by connecting peptide segements X_n, X_m, and X_p, respectively, wherein n=0-5, m=0-5 and p=3-8 amino acid residues, respectively, and the conjugation moiety F_L is a cysteine residue. In a further embodiment the functional moieties F_1, F_2, and F_3 on the ligand heads have been selected by bacterial phage display for optimal receptor binding. In an embodiment the functional moieties and connecting peptide segments of an active oligopeptide ligand head are replaced by chemical mimetics. In another embodiment an appropriate chemical scaffold of connecting segments has been designed to comprise (present) functional moieties F_1, F_2, and F_3, which have been selected by combinatorial chemistry for optimal receptor binding from a library of chemical moieties complementary to receptor-binding sites on the surface of SCF. In another embodiment the linker comprises an organic polymer having two ends capped at each end by a reactive capping moiety, F_c, which react covalently with the conjugation moiety, F_L, on the ligand head. In a further embodiment the organic polymer is polyethyleneglycol (PEG) comprising the structure $H[OCH_2CH_2]_nOH$, wherein n is 10-20. In another embodiment the capping moiety, F_c, is a thiol-reactive group such as

Sub C12 Cont-8

N-ethyl maleimide. In an embodiment the conjugating moiety, F_L, is a thiol containing group such as cysteine.

Sub C13

This invention provides a method of treating a subject comprising administration of a compound designed by the above described method. In an embodiment the subject has a blood disorder. In a further embodiment the blood disorder is anemia or immunodeficiency. In an embodiment the compound is administered orally or any other routes. In an embodiment the compound is an isolated SCF analog. In another embodiment the compound comprises an isolated SCF analog, whose alteration site is a receptor binding site on the surface of the altered SCF molecule. In another embodiment of the method the composition comprises a double-headed receptor SCF ligand analog having the structure set forth in Figure 10A. In an embodiment each ligand head of the double-headed SCF ligand analog is an oligopeptide having the structure set forth in Figure 10B. In another embodiment the ooligopeptide comprises a sequence, wherein functional moiety F_1 corresponds to a segment of amino acid residues from within N-terminal residues 1-10 of SCF, functional moiety F_2 corresponds to a segment of amino acid residues from within residues 79-95 of SCF, and functional moiety F_3 corresponds to a segment of amino acid residues located within three amino acid residues of amino acid residue 127, wherein F_1, F_2, and F_3 are connected by connecting peptide segements X_n, X_m, and X_p, respectively, wherein n=0-5, m=0-5 and p=3-8 amino acid residues, respectively, and the conjugation moiety F_L is a cysteine residue. In a further embodiment the functional moieties F_1, F_2, and F_3

Sub C13 cont'd

on the ligand heads have been selected by bacterial phage display for optimal receptor binding. In an embodiment the functional moieties and connecting peptide segments of an active oligopeptide ligand head are replaced by chemical mimetics. In another embodiment an appropriate chemical scaffold of connecting segments has been designed to comprise (present) functional moieties F_1, F_2, and F_3, which have been selected by combinatorial chemistry for optimal receptor binding from a library of chemical moieties complementary to receptor-binding sites on the surface of SCF. In another embodiment the linker comprises an organic polymer having two ends capped at each end by a reactive capping moiety, F_c, which react covalently with the conjugation moiety, F_L, on the ligand head. In a further embodiment the organic polymer is polyethyleneglycol (PEG) comprising the structure $H[OCH_2CH_2]_nOH$, wherein n is 10-20. In another embodiment the capping moiety, F_c, is a thiol-reactive group such as N-ethyl maleimide. In an embodiment the conjugating moiety, F_L, is a thiol containing group such as cysteine.

This invention provides a method of stimulating the production of hematopoietic cells in a subject comprising administering an isolated stem cell factor (SCF) analog. In an embodiment isolated stem cell factor (SCF) analog is prepared by the method of claim 1 or designed by the above described method. In another embodiment the administration is oral or any other route. In an embodiment the isolated SCF analog has an alteration in at least one atom of the atomic coordinates of the crystal structure as set forth in Figure 8. In another

embodiment the isolated SCF analog comprises amino acid residues of native or recombinant SCF1-165 or amino acid residues of a recombinant selenomethionyl SCF1-141. In an embodiment of this method the isolated SCF analog, comprises an isolated altered SCF molecule, whose alteration site is a receptor binding site on the surface of the altered SCF molecule. In another embodiment of the above-described the compound comprises an isolated SCF analog, whose alteration site is a receptor-binding site on the surface of the altered SCF molecule. In another embodiment of said method the composition comprises a double-headed receptor SCF ligand analog having the structure set forth in Figure 10A. In an embodiment each ligand head of the double-headed SCF ligand analog is an oligopeptide having the structure set forth in Figure 10B. In another embodiment the ooligopeptide comprises a sequence, wherein functional moiety F_1 corresponds to a segment of amino acid residues from within N-terminal residues 1-10 of SCF, functional moiety F_2 corresponds to a segment of amino acid residues from within residues 79-95 of SCF, and functional moiety F_3 corresponds to a segment of amino acid residues located within three amino acid residues of amino acid residue 127, wherein F_1, F_2, and F_3 are connected by connecting peptide segements X_n, X_m, and X_p, respectively, wherein n=0-5, m=0-5 and p=3-8 amino acid residues, respectively, and the conjugation moiety F_L is a cysteine residue. In a further embodiment the functional moieties F_1, F_2, and F_3 on the ligand heads have been selected by bacterial phage display for optimal receptor binding. In an embodiment the functional moieties

and connecting peptide segments of an active oligopeptide ligand head are replaced by chemical mimetics. In another embodiment an appropriate chemical scaffold of connecting segments has been designed to comprise (present) functional moieties F_1, F_2, and F_3 which have been selected by combinatorial chemistry for optimal receptor binding from a library of chemical moieties complementary to receptor-binding sites on the surface of SCF. In another embodiment the linker comprises an organic polymer having two ends capped at each end by a reactive capping moiety, F_c, which react covalently with the conjugation moiety, F_L, on the ligand head. In a further embodiment the organic polymer is polyethyleneglycol (PEG) comprising the structure $H[OCH_2CH_2]_nOH$, wherein n is 10-20. In another embodiment the capping moiety, F_c, is a thiol-reactive group such as N-ethyl maleimide. In an embodiment the conjugating moiety, F_L, is a thiol containing group such as cysteine.

Sub C14

This invention provides an isolated stem cell factor (SCF) molecule, which is an altered SCF, comprising any portion of amino acids 1-165 of a human SCF polypeptide, optionally comprising an N-terminal methionine before amino acid residue 1, wherein the polypeptide has an amino acid sequence portion of SCF capable of binding to the SCF receptor. In an embodiment of the altered isolated stem cell factor molecule an alteration is selected from the group consisting of deletion, insertion and substitution of at least one amino acid residue from the naturally occurring amino acid sequence of SCF.

Sub C15

In a further embodiment an alteration is a truncated SCF comprising amino acids 1-141 of a human SCF polypeptide, optionally comprising an N-terminal methionine before amino acid residue 1, E. In another embodiment the three-dimensional structure is altered from the atomic coordinates are set forth in Figure 8. In yet another embodiment the electron density distribution map is altered from the atomic coordinates are set forth in Figures 1A, 1B, or 1C. In a still further embodiment the substitution of at least one amino acid residue is selected from the group consisting of SCF(Y26C) disulfide-linked dimer, SCF(D25C), SCF(K62C), SCF(K78N, N81K), SCF(R117A, I118A), SCF(E92A, S95A), and SCF(D124A, K127D). In another embodiment the overall three-dimensional conformation of the stem cell factor molecule has an altered three-dimensional structure of the αC-β2 loop.

Sub C16

This invention provides a pharmaceutical composition comprising the above described altered isolated SCF molecule and a pharmaceutically acceptable carrier. In an embodiment the altered SCF molecule molecule is a hybrid molecule of the altered stem cell factor molecule and a second protein or fragment thereof. As used herein, an SCF hybrid molecule is defined as a molecule wherein analog SCF is combined with with part or all of another protein such as another cytokine or another protein, which for example, effects signal transduction via entry through the cell through a SCF-SCF receptor transport mechanism. In an embodiment the alteration of the αC-β2 loop is a change in length of the amino acid sequence of

Sub C1 cont'd

the αC-β2 loop by a deletion or an insertion of at least one amino acid residue or a change in at least one amino acid residue from the naturally occurring amino acid residue(s) of the αC-β2 loop. In another embodiment the change in said at least one amino acid residue from the naturally occurring amino acid residue(s) is selected from the group consisting of SCF(Y26C) disulfide-linked dimer, SCF(D25C), SCF(K62C), SCF(K78N, N81K), SCF(R117A, I118A), SCF(E92A, S95A), and SCF(D124A, K127D).

Generally, for design of drugs as described in the above-described methods, certain changes are known to have certain structural effects. For example, deleting one cysteine could result in the unfolding of a molecule which is, in its unaltered state, is normally folded via a disulfide bridge. There are other known methods for adding, deleting or substituting amino acids in order to change the function of a protein.

The atomic coordinates may be determined in the above-described method by multiwave anomalous diffraction (MAD) measurements, but is not limited htereto, since any means determined suitable by one of skill in the art may also be used.

This invention will be better understood from the Experimental Details which follow. However, one skilled in the art will readily appreciate that the specific methods and results discussed are merely illustrative of the invention as described more fully in the claims which follow thereafter.

EXPERIMENTAL DETAILS

MATERIALS AND METHODS

SCF Expression, purification and analyses

Sub C17

Human SCF$^{1-141}$ was expressed recombinantly in *E. coli* as described previously (Langley *et al.*, 1994). For expression of SeMet SCF$^{1-141}$, the expression vector was transfected into the methionine auxotrophic *E. coli* strain FM5. Fermentation was carried out at 30°C in 8 liters of minimal medium consisting of ammonium sulfate (10 g/liter), glucose (5 g/liter), methionine (0.125 g/liter), phosphate salts, magnesium, citric acid, trace metals, and vitamins. When an OD$_{600}$ of 3-5 was reached, a feed medium was added that consisted of the following components in a total volume of 1 liter: 100 g of ammonium sulfate, 450 g of glucose, 2 g of methionine, magnesium, trace metals, and vitamins. At an OD$_{600}$ of 12.4, induction medium (one liter containing 100 g of ammonium sulfate, 300 g of glucose, and 1 g of selenomethionine) was added and fermentation proceeded at 30°C. Five hours later (at an OD$_{600}$ of approximately 16), the temperature was raised to 42°C to induce SCF expression and additional selenomethionine (1 g) was added. Cells were harvested 4 hours after the temperature shift (OD$_{600}$ of approximately 16). SeMet SCF$^{1-141}$ expression was estimated as 0.5 g/liter. Both SCF$^{1-141}$ and SeMetSCF$^{1-141}$ were purified with minor modifications to previously described procedures (Langley et al., 1992, 1994). Both retain the initiating

Sub C17 Cont'd

methionine (or SeMet) residue [position (-1)] (Langley et al., 1994). N-terminal amino acid sequencing was performed as described (Lu et al., 1991). About 90% SeMet was present in SeMetSCF$^{1-141}$ at each of the Met positions, based on amino acid analysis and N-terminal sequencing results (i.e. lack of recovery of Met residues for SeMetSCF$^{1-141}$ in comparison with SCF$^{1-141}$; data not shown).

Crystallization

Crystals were obtained by the use of hanging drop vapor diffusion method under aerobic conditions. The initial crystals were grown by mixing 1 μl of protein solution [44 mg/ml for SCF$^{1-141}$ or 38 mg/ml for SeMet SCF$^{1-141}$) in 10 mM sodium phosphate pH 6.5, 80 mM NaCl] with 1 μl crystallization reservoir solution. The crystallization reservoir solution included 25% (w/w) PEG 400, 240 mM $CaCl_2$, 100 mM HEPES pH 7.4 for SCF$^{1-141}$, and 22% PEG400, 220 mM $CaCl_2$, 100 mM HEPES pH 7.2 and 5-10 mM dithiothreitol (DTT) for SeMetSCF$^{1-141}$. Crystallization trays were incubated at 20° C and crystals reached full size in approximately 3 days with typical dimensions of 0.5 x 0.2 x 0.2 mm. Microseeding and lower concentrations of DTT solution (2 mM) wre needed to reproduce SeMetSCF$^{1-141}$ crystals subsequently. An extant SeMetSCF$^{1-141}$ crystal was washed with its reservoir solution and then crushed to produce microseeds, which were stored in 50 μl of a stabilizing solution of 32% (w/w)PEG400, 260 mM $CaCl_2$, 100 mM HEPES (pH 7.4) at room temperature. For microseeding experiments, the seed stock was diluted by 10-10,000-fold with crystallization reservoir solution. A 1 μl aliquot of this prepared precipitant was mixed with 1 μl of the

protein solution to make the droplet. The crystal for MAD phasing was grown from a crystallization reservoir solution containing 2 mM DTT concentration.

Diffraction measurements

X-ray diffraction data from SCF^{1-141} crystals were recorded on two Hamlin-Xuong area detectors at 293K at a home source. The data were integrated using the UCSD software package and scaled using AGROVATA and ROTAVATA as implemented in CCP4 suite (CCP4, 1994). The MAD experiments for $SeMetSCF^{1-141}$ were conducted at the X4A synchrotron beam line of Brookhaven National Laboratory using Fuji image plates. A single crystal was frozen at 110K using paratone-N (Exxon) as a cryoprotectant. The MAD data were collected at four wavelengths (before the edge, at the SeK edge, at the peak and after the peak) in oscillations of 1.3-1.5° without overlap. The $SeMetSCF^{1-141}$ crystal was oriented such that *b*-axis was parallel to the oscillation axis and a mirror geometry was used during data collection. The MAD data were processed using DENZO and Scalepack (Otwinowski, 1993; Gewirth, 1995)(Table I).

Table I. MAD data collection and phasing statistics.

Data collection (25 - 2.0 Å)[a]

Wavelength(Å)	Unique reflections	Completeness (%)	Signal(I/σ)	R_{sym}(%)
λ1=0.9919 (pre-edge)	65,810	95.1	18.4	6.7
λ2=0.9793 (inflection)	65,759	95.0	16.7	5.8
λ3=0.9791 (peak)	65,665	94.9	15.2	6.7
λ4=0.9686 (remote)	65,689	94.9	16.0	5.6

Anomalous diffraction ratios (20 - 2.6 Å)[b]

	λ1	λ2	λ3	λ4	f' (e)	f" (e)
λ1	0.035 (0.030)	0.051	0.042	0.035	-4.0	-0.5
λ2		0.052 (0.029)	0.033	0.051	-10.3	3.8
λ3			0.070 (0.031)	0.041	-8.1	5.6
λ4				0.055 (0.030)	-3.9	3.8

MAD phasing (25 - 2.6 Å)[c]

$R(°|F_T|) = 0.044$ $R(°|F_A|)=0.39$ $<\Delta(\Delta\Phi)> = 41.6°$ $<\sigma(\Delta\Phi)> = 18.7°$

Table I continued. MAD data collection and phasing statistics.

[a] Unique reflections are determined by point group 222 (not mmm) to distinguish Bijvoet-related reflections. Rsym = 100 x $\Sigma_{hkl}\ \Sigma_i\ |I_i - \langle I\rangle| / \Sigma_{hkl}\Sigma_i I_i$, where I_i is the ith measurement of reflection hkl and <I> is the weighted mean of all measurements of I.

[b] Anomalous diffraction ratios = $\langle\Delta|F|^2\rangle^{½}/\langle|F|^2\rangle^{½}$, where $\Delta|F|$ is the absolute value of the Bijvoet (diagonal elements) or dispersive difference (off-diagonal elements), respectively. Values in parentheses are for centric data.

[c] $R = \Sigma_{hkl}\ \Sigma_i\ |\ |F_i| - \langle F\rangle\ |\ /\ \Sigma\ |F|$. $^\circ F_T$ is the structure factor due to normal scattering from all the atoms. $^\circ F_A$ is the structure factor due to normal scattering from the anomalous scatterers only, and $\Delta\Phi$ is the phase difference between $^\circ F_T$ and $^\circ F_A$. $\Delta(\Delta\Phi)$ is the difference between two independent determinations of $\Delta\Phi$.

Molecular replacement attempts

Structure determination by the molecular replacement method was attempted for the home source data set. The MERLOT (Fitzgerald, 1988) and AmoRe (CCP4, 1994) programs were used with various four-helix bundle structures as search models, and a good rotation solution was obtained. The rotation solution agreed well with the orientation of helical bundles (approximately along the *b*-axis of unit cell) that was deduced from native Patterson maps. Dissimilarities among the helical cytokines and the multiplicity of subunits (four) hampered detection of any significant translational function peaks.

Phase evaluation

The processed MAD data were passed through the MADSYS programs (Hendrickson, 1985). Algebraic and probabilistic MAD phasing procedures (Hendrickson, 1985; Pahler *et al.*, 1990) were applied for phase determination (Table II). Selenium sites were located by HASSP program (CCP4, 1994) in F_A Patterson and difference Fourier maps and refined by MADSYS programs. The choice of enantiomer was determined by comparison of the electron density maps computed from the two enantiomorphic selenium structures to maximum Bragg spacings of 2.6 Å. The phases were improved by 4-fold non-crystallographic symmetry (NCS) averaging. The rotation-translation matrices of the NCS axes were determined by TOSS (Hendrickson, 1979) from the selenium sites and subsequently refined by LSQRHO (W.A. Hendrickson, unpublished) and RAVE (Kleywegt and Jones, 1994), and the averaging procedure by DM (CCP4, 1994).

Model building and refinement

The initial model of SeMetSCF$^{1-141}$ was built into the averaged map at 2.3 Å by using program O (Jones *et al.*, 1991). The model includes 98 core residues for each of the four molecules in an asymmetric unit. The remote wavelength after the SeK peak was used for the refinement with the Bijvoet difference applied to Se scattering factors. The *R*-value for this model, before any refinement, was 42.1% in the resolution range of 10.0 - 2.3 Å. NCS restraints were applied during the initial rounds of refinements. After several iterations of least square and simulated annealing refinement with X-PLOR (Brunger *et al.*, 1987) and manual rebuilding against SIGMAA (Read, 1986) and $2|F_o|-|F_c|$ maps, the crystallographic *R*-value is 19.9% for the current model (Table III). The sites of Ca^{2+} ions, a component of the crystallization medium, were located from a Bijvoet difference Patterson map at the remote wavelength before the SeK edge. The SCF$^{1-141}$ model was obtained by subjecting the refined SeMetSCF$^{1-141}$ model to refinement against the area-detector data set from the SCF$^{1-141}$ crystal using the XPLOR program (Brünger *et al.*, 1987). The atomic coordinates have been deposited in the Brookhaven Protein Data Bank with accession code 1scf.

Table II. Lattice and Refinement Statistics

	SeMeSCF$^{1-141}$ (λ4)	Native
Lattice		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell constants (a,b,c) (Å)	71.8, 82.6, 88.2	73.0, 84.7, 88.8
Z_a[a]	4	4
Refinement[b]		
Resolution range (Å)	20.0 - 2.2	8.0 - 3.3
Completeness (%)	96.6	98.6
Unique reflections[c]	49851	7990
R-value[d] ($\lvert F\rvert>2\sigma$) (%)	19.9	20.8
R_{free}[e] (%)	24.2	27.3
R_{sym}[f] (%)	5.6	15.2
Model parameter		
Total non-H atoms	3804	3502
Total residues	448	447
Total water molecules	264	0
Total metal ions	3	0
rms bond length/angle	0.016/2.5°	0.017 / 3.0°
Average *B*-factor (Å^2)	32.1	18.7
main-chain rms *B* (bond, angle) (Å^2)	1.2/1.6	1.9/2.2
side-chain rms *B* (bond, angle) (Å^2)	2.1/2.4	3.0/3.3

[a]Z_a: number of molecules in the asymmetric unit.

[b]The reflection data higher that the resolution range were not included in the refinement due to poor R_{sym} in these resolution shells.

[c]Unique reflections are determined by point group 222 for the SeMetSCF$^{1-141}$ dataset to distinguish Bijvoet-related reflections and by point group mmm for native dataset.

[d]R-value = $\Sigma_{hkl}\lvert\lvert F_o\rvert - \lvert F_c\rvert\rvert / \Sigma_{hkl}\lvert F_o\rvert$.

Table II. continued Lattice and Refinement Statistics

[e]A subset of the data (6%) was excluded from the refinement and used for the free *R*-value calculation.
[f] R_{sym} for $SeMetSCF^{1-141}$ data set was calculated in the resolution range of 25-2.2 Å and for the SCF^{1-141} data set in the resolution range of 13-3.3 Å.

Structure analysis

Solvent accessibilities were defined as compared with the corresponding Gly-X-Gly peptide (Shrake and Rupley, 1973) as calculated by XPLOR (Brunger *et al.*, 1987). Structural superimpositions were performed based on a-carbon atoms alone. The coordinates were taken from the Brookhaven Data Bank with entry codes: M-CSF, 1hmc (Pandit *et al.*, 1992); IL-4, 1rcb (Wlodawer *et al.*, 1992); GM-CSF, 1gmf (Diederichs *et al.*, 1991) IL-2, 3ink (McKay, 1992); IL-5, 1hul (Milburn *et al.*, 1993). Initial segments of equivalence between two structures were defined according to equivalent secondary structure elements. These structures were then superimposed using program TOSS (Hendrickson, 1979) and the number of equivalent atoms were extended using Lsq_imp command in program O (Jones *et al.*, 1991). A cutoff distance of 3.0 Å and at least three residues in a consecutive fragment were used as the criteria of defining equivalent atom sets. Different initial equivalent segments did give different results in the structural alignment, as Rozwarski *et al* observed in their study (Rozwarski *et al.*, 1994). In this study, several initial sets of equivalent segments for each alignment were tried and the one that generated in the greatest number of equivalent atoms after the Lsq_imp extension was retained.

RESULTS AND DISCUSSION

Structure determination

Both native and selenomethionyl (SeMet) human SCF^{1-141} were expressed as recombinant proteins in *E. coli* (Langley *et al.*, 1994). Crystals grew in space group $P2_12_12_1$ with four SCF subunits and 39% solvent in the asymmetric unit. The attempts to solve the crystal structure of SCF^{1-141} by molecular replacement from other cytokine models gave good rotation solutions, but no significant translation function peaks. Experimental phases for $SeMetSCF^{1-141}$ were then evaluated in a multiwavelength anomalous diffraction (MAD) experiment. Four-wavelength data were measured from a single, frozen $SeMetSCF^{1-141}$ crystal and analyzed with MADSYS (Hendrickson, 1985). Twelve selenium sites were found in four congruent sets that proved to be associated with the respective SCF subunits in the crystal. A MAD-phased electron-density map was calculated at 2.3Å resolution (**Figure 1A**) and improved by molecular averaging (**Figure 1B**) and refinement (**Figure 1C**).

An atomic model was fitted to the experimental maps and refined at 2.2Å resolution to an *R*-value of 0.199 ($|F| > 2\sigma$) with stereochemical ideality typified by the r.m.s. deviation from bond ideality of 0.016Å. There are no residues in energetically disfavored regions of the Ramachandran plot. This model for $SeMetSCF^{1-141}$ has 3804 non-hydrogen atoms from 448 amino acid residues, 264 water molecules, three Ca^{2+} ions and one polyethylene glycol (PEG) moiety. All four polypeptide chains

(designated A, B, C, and D) are sufficiently disordered before residue 11 to preclude modeling of this portion, and none of them is fully ordered through to the end. Specifically, A92-103, B130-136, B139-141, C92-103, C127-141, and D91-103 and D128-141 are all disordered. This disorder is such that, of the eight disulfide bridges, only two are seen. To test whether the reducing agent used to crystallize SeMetSCF$^{1-141}$ (see Materials and Methods) might have broken these bonds and caused the disorder, the native SCF$^{1-141}$ structure which was crystallized without reducing agent was also refined. The two crystals are nearly isomorphous (differences are due to temperature at data collection), and the two structures show the same pattern of order-disorder.

Structure of SCF

The four independent SCF subunits in the crystal are similar but distinctive, and identification of the AB and CD pairs as the molecular dimers is unmistakable. None of the SCF monomer copies is complete, but each flexible portion except for the N-terminus is stabilized by lattice contacts to another monomer. Thus, through the combination of chains A and B there are images for all but residues 1-10, and the position of Cys89 to which Cys 4 must bridge, determines the approximate course of this disordered segment. The overall structure of this composite SCF dimer is shown in **Figure 2A** and the C_α backbone for the actual AB dimer is drawn in stereo in **Figure 2B**. Topologically, SCF structure is similar to other short-chain helical cytokines (Rozwarski *et al.*,

1994) with a core of four helices (αA, αB, αC and αD) and two beta strands, β1 between αA and αB and β2 between αC and αD. Apart from the tight β2-αD connection, however, the segments outside these core elements are unique in conformation if not in length. In particular, there is an additional one-turn helix, αB', between β1 and αB, there is an exceptional hairpin loop between αB and αC at the dimer interface, and there is another extra one-turn helix, αD', in the C-terminal extension. The bounds of secondary-structure elements are given in **Figure 3**.

The core SCF dimer has its subunits arranged in a head-to-head manner with the opposed four-helix bundle axes nearly coincident (**Figure 2**). This gives the molecule an elongated shape, ~ 85Å x 30Å x 20Å. Approximately 855 Å^2 of surface area is buried from each protomer into the dimer interface. The interface is dominated by contacts from the C-terminal end of αA and the αA-β1 connection of one monomer to the αB-αC loop of the other monomer (**Figure 2**), and the reciprocal pair is related by an approximate dyad axis of symmetry. The actual symmetry operators have rotational and translational components of 176.3° and 0.33°, respectively, for the AB dimer and 177.4° and 0.04° Å for the CD dimer. The two dimers thereby deviate significantly and similarly (with A matched to C and B matched to D) from true 2-fold symmetry. Nevertheless, since interatomic contacts at the interface are symmetric, it is presumed that thses deviations reflect flexibility rather than inherent asymmetry.

Then the r.m.s. deviation for the C_α positions in common between the two dimers is 0.80 Å (208 C_α atoms) is comparable to that of pairwise comparisons among the four independent molecules (from 0.57 Å to 0.94 Å for 103 C_α atoms). If D alone is superimposed onto B, a rotation of 2.1° brings A and C into optimal superposition. In the contrary match-up, with D onto A, a rotation of 6.7° is needed to superimpose B and C.

The crystal structure is compatible with solution biochemistry. Consistent with the relative rates of in vitro oxidation of methionyl residues (Hsu et al. 1996), Met36 and Met48 are buried in the hydrophobic core whereas Met27 is solvent accessible. Furthermore, as predicted on the basis of fluorescence spectroscopy studies (Arakawa et al., 1991), Trp41 is buried within the hydrophobic core.

Natural SCF and Chinese hamster ovary(CHO) cell-expressed recombinant SCF are heavily glycosylated by both N-linked and O-linked carbohydrates. All four potential N-linked sites are in the SCF^{1-165} are in the SCF^{1-141} portion that has been crystallized (Langley et al., 1992; Lu *et al.*, 1992). Although the recombinant proteins expressed in bacteria are non-glycosylated, both human and rat SCF expressed in *E. coli* and then refolded *in vitro* have native structures, as judged by biophysical methods and *in vitro* biopotency assays (Arakawa et al., 1991; Langley et al., 1992) . The crystal structure of the recombinant SCF in this study is compatible with the glycosylation pattern found for SCF expressed from mammalian cells.

Thus, the potential site at Asn72, which is unglycosylated in both human and rat natural SCF expressed from mammaliam cells, is buried in the dimer interface, whereas the site at Asn120, which is fully glycosylated in both species, is accessible in the atomic model. Other sites (Asn65 in both human and rat, human Asn93 and rat Asn109) are glycosylated in some molecules but not others. These sites are also accessible in the atomic model. Asn93 is located in the highly flexible region between αC and β2, and its side chain is disordered.

Although natural SCF is a noncovalently associated dimer, recombinant human SCF produced in *E. coli* can fold alternatively *in vitro* into a covalently-linked dimer. These dimers have Cys4-Cys89' and Cys43-Cys138' intermolecular disulfide bonds (Lu *et al.*, 1996). The disulfide-linked and natural non-covalently associated SCF dimers are similar with regard to biochemical and biophysical properties, biopotency and receptor-binding affinity. The disulfide-linked SCF is also biologically active with higher biopotency in supporting growth of hematopoietic cell line and stimulating hematopoietic cell colony formation but slightly lower binding affinity to c-Kit than the noncovalently associated dimer. It was proposed that the disulfide-linked dimer arises from a double-swap of αA and αD helices between the monomers (Lu *et al.*, 1996). The crystal structure of SCF, however, suggests that a single-swap at the αB-αC loop near residue 68 is more likely.

Comparison with other short-chain helical cytokines

Sub 33

Although SCF has the characteristic features of short-chain helical cytokines, as among other members, both sequence and structure are highly divergent. If anything, SCF resembles the others less than they resemble one another (**Table III**). The comparison in this study of SCF with other short-chain helical cytokine structures [granulocyte-macrophage colony-stimulsting fasctor (GM-CSF) (Diederichs et al., 1991), , M-CSF (Pandit e al., 1992), interleukin (IL)-2 (McKay, 1992), IL-4 (Wlodaver et al., 1992) , and IL-5 (Milburn et al., 1993)] shows greatest structural similarity with M-CSF or IL-4, but even here fewer than half of the residues can be superimposed (**Table III**). Sequence similarities are essentially random. A structure-based sequence alignment (**Figure 3**) of SCF with other short-chain helical cytokines has pairwise identities ranging from 6.7% to 18.8% (**Table III**) and not even a single residue in SCF is conserved in all the others. Moreover, the best alignment presented in Figure 3 is only valid for the specified criteria herein, and it differs somewhat from that given by Rozwarski et al. (Rozwarski *et al.*, 1994). Indeed, because of variability in the core structures in this divergent superfamily, a self-consistent pairwise alignment of the family members has not been able to be achieved. Nevertheless, the core elements are remarkably similar in structure.

Table III. Structural and sequence comparisons of short-chain helical cytokines.

	SCF	M-CSF	IL-4	GM-CSF	IL-2	IL-5
SCF		14.1 (13.0)	12.7 (12.3)	12.5 (23.5)	18.8 (16.4)	6.7 (21.1)
M-CSF	64 (1.755)		14.8 (18.9)	13.8 (18.3)	17.5 (17.1)	10.5 (18.6)
Il-4	63 (1.578)	54 (1.820)		26.6 (25.0)	14.5 (22.2)	18.9 (18.9)
GM-CSF	48 (1.632)	58 (1.814)	64 (1.559)		9.8 (26.0)	20.4 (14.7)
IL-2	48 (1.700)	57 (1.581)	69 (1.330)	61 (1.482)		14.5 (22.2)
IL-5	45 (1.695)	38 (1.721)	53 (1.324)	49 (1.334)	62 (1.371)	

Structural comparisons and sequence comparisons between the short-chain helical cytokines are given in the lower and upper triangles, respectively. Structural comparisons are given as the maximum number of equivalent α-carbon atoms between two short-chain helical cytokines, and the r.m.s. deviation (Å), (in parentheses). Sequence comparisons are given as the percentage of sequence identity from sequence alignment based on structural superimposition, and that based on the sequence alignment from BESTFIT program of the GCG package (in

Table III. continued Structural and sequence comparisons of short-chain helical cytokines.

parentheses). The latter alignment is based only on maximizing the percentage of identity, similarities and length of the matching sequences, and the sequences submitted to the BESTFIT program were restricted within the region as defined in the PDB files, including the disordered residues. With the advantage of the relatively large number of independent data points (15 pairs), the correlation between sequence similarity and structural deviation was analyzed. Without any restriction of structural alignment, the correlation coefficient (C) between structural deviation and sequence identity is -0.21 and the student's t probability (P) is 0.44, suggesting little correlation between a specific sequence and the tertiary fold. With the restriction of structural alignment, however, C is-0.30 and P 0.28, indicating that the structure-based sequence identity and structural deviation are weakly connected (as also observed in another highly diverged protein family, hemoglobin; Aronson *et al.*, 1994).

Core portions aside, SCF differs markedly from other short-chain helical cytokines, as indeed they differ from one another (Figure 3; Rozwarski *et al.*, 1994)). First, helix αA of SCF is unusually shortened at its N-terminus. Its disordered extension must deviate toward αC, as in M-CSF but not in the others, by virtue of the Cys4-Cys89' disulfide bridge in common with M-CSF. Secondly, the conformation of the αA-β1 connection is distinctive as required for the dimer interface, and the β1-αB connection uniquely has αB'. Again at the dimer interface, the αB-αC loop extends out distinctively along the dyad axis. Thirdly, the unusually long αC-β2 loop of SCF is both highly flexible (only one ordered copy) and with a path of its own when ordered. Finally, the C-terminal extension after αD compares only to that of M-CSF, and then only in its general direction of exit out past αB and the β- strands.

Among the short-chain helical cytokines, SCF is most closely related to M-CSF. These two have similarities in gene structure, alternative splicing, proteolytic maturation, disulfide bridging, dimer assembly, and receptor type (these similarities also extend to the Flt-3igand; Lyman and Jacobsen, 1998). Despite negligible sequence identity, an alignment and secondary structure prediction prompted by these relationships (Bazan, 1991) fits the actual

structure amazingly well, except for shifts in αB and in the αC-β2 loop. Here reality confounds logic; unexpectedly, comparable glycosylation sites (Asn120 in SCF and Asn122 in M-CSF) are displaced by one helical turn and comparable disulfide bridges (Cys43-Cys138' in SCF and Cys48-Cys139' in M-CSF) are not superimposible structurally (**Figure 4**).

Both were roughly correct in secondary-structure prediction for helices αA and αC, but substantial misplacements were made for helices αB and αD and strand β2. In the study of Rozwarski et al. (Rozwarski *et al.*, 1994), the alignment for αB is incorrect by a shift of 14 residues and that for β2 and αD by a shift of 7 residues. Bazan's earlier sequence alignment (Bazan, 1991) fits to the structural alignment herein amazingly well, except for a shift of one residue for αB and a three-residue gap in the αC-β2 loop.

Comparison with other cytokine dimers

Helical cytokines dimerize in various ways (Sprang and Bazan, 1993). Among the five dimeric helical cytokines for which crystal structures have been described [M-CSF, IL-5, ciliary neurotrophic factor (CNTF), interferon-γ (IFN-γ) and IL-10], only IFN-γ and IL-10 are similar dimers. These latter two have a 'tip-to-tip' packing with helix axes approximately perpendicular. Otherwise, the only salient

feature in common is having the subunits oriented with bundle axes aligned in parallel and helix dipoles positioned to compensate. There is 'head-to-head' packing of the four-helix bundles in M-CSF, 'tail-to-tail' packing in IL-5, and 'side-to-side' packing in CNTF. Moreover, IFN-γ, IL-10 and IL-5 are all interdigitated dimers with helices swapped between subunits. Thus, although SCF relates most closely to M-CSF, the dimer structure could not be deduced readily beforehand.

SCF in keeping with its relationship to M-CSF, is a non-interdigitated 'head-to-head' dimer (**Figure 4**). The two interfaces between promoters are completely different, however. One αA-β1 loop of M-CSF is situated between the αA-β1 and αB-αC loops of the other protomer, whereas in SCF each αA-β1 loop interacts only with αB-αC loop of the partner. This staggered mode of M-CSF dimerization (**Figure 4B**) is dictated by the position of the Cys31-Cys31' intermolecular disulfide bond in M-CSF. The dyad axes are similarly oriented in the two cases (perpendicular to the bundle axis and parallel to the αA-αD and αB-αC helix planes), but whereas the dyad axis in SCF nearly intersects the bundle axis, that in M-CSF is offset toward the αA-αD helix pair (Figure 4). Thus, when one protomer of an SCF dimer is superimposed onto one from M-CSF, the superimposition of the two mates requires a translation of 3.8 Å but a rotation of only 4.7°.

Location of the binding site for the receptor Kit

SCF binds with high affinity (nM range) to its receptor (Philo *et al.*, 1996; Broudy, 1997)). Various structure-function studies and analyses help to define residues of SCF that may be involved in this binding. These studies include mutagenesis experiments, immunochemical mapping, comparative analyses of inter-species ligand-receptor interactions, and analyses of glycosylation. Residues thereby implicated in receptor binding can then be mapped onto the surface of SCF as defined by the crystal structure. Although a precise definition of the receptor-binding site on SCF will require direct structural information on the complex of SCF with the Kit receptor, this mapping of the binding site provides a crude picture that is useful when coupled with information on Kit and related receptors.

Sub C18

From studies of truncation and point mutants, Langley *et al* (1994) demonstrated that the N-terminal residues 1-4 and 1-10 and the Cys4-Cys89 disulfide bond are required for receptor binding and bioactivity, and that the Cys43-Cys138 disulfide bond and C-terminal residues past 127 are not required for receptor binding but may have some roles in cell proliferation activity. Moreover, alterations at Asn10 and Asn11 brought about by chemical isomerization or by mutagenesis have positive or negative effects depending on the substitution (Hsu *et al.*, 1998). A quadruple

Sub C18 cont'd

mutant of SCF (Arg121Asn, Asp124Asn, Lys127Asp and Asp128Lys) was found to be defective in bioactivity (Matous *et al.*, 1996). The molecular cause of this deficiency may be specific to Lys127 or due to indirect electrostatic effects. Arg121 and Asp124 are adjacent to the main N-linked glycosylation site, which is not involved in binding (see infra), and Asp128 is absent in the 1-127 truncation mutant that retains full receptor-binding activity (Langley *et al.*, 1994). Moreover, a study of human-murine SCF chimeras narrowed the important receptor recognition epitopes to within residues 1 to 35 and 79 to 97 (Matous *et al.*, 1996), and the epitope of a neutralizing antibody was mapped to the region of residues 60-95 (Mendiaz et al., 1996) and 79-97 (Matous et al., 1996).

Although SCF molecules from different mammalian species are very similar (>75% identity), there are substantial differences in inter-species receptor activation. Human SCF activates murine Kit very poorly, rodent SCF has only slightly lower potency than human SCF in binding/activating human Kit (Martin *et al.*, 1990; Lev *et al.*, 1992), and canine SCF activates human Kit slightly better than human SCF does itself (K.E. Lang, unpublished data). It is likely that the receptor-binding regions involve residues that are different between man and mouse but conserved between man and dog. These residues can be classified into five groups in the sequence (**Figure 5**). Most residues in group III are buried

and those in group II are close to the dimer interface. The residues in groups III (45-58) are buried and those in group II (24-34) are close to the dimer interface. The results in groups I (1-15), IV (80-117) and and V (130-140) are more likely to be involved in direct receptor binding.

The heavy glycosylation of natural and CHO cell-derived recombinant SCFs sheds light on the question whether residues in vicinity of αD, the equivalent of the major receptor binding site in GH, are involved in receptor binding. Human SCF expressed in CHO cells is approximately 30% by weight (Arakawa *et al.*, 1991) The main glycosylation site is at Asn120 (Langley et al., 1992). Glycosylation at this site, which is near the center of the αD helix, does not appear to influence biological activity; therefore, the area around this residue cannot be involved in receptor binding. Glycosylation of human SCF at either Asn65 or Asn93 lowers the biological activity approximately 10-fold; therefore, these residues may be near but not directly at the binding site.

Taken together, these observations indicate that the receptor-binding site may include residues from the first few N-terminal residues, the 79-95 region (mainly located on αC helix) and the C-terminal end of αD (around 127). These regions are contiguous on the SCF surface in the atomic model provided herein. The putative receptor-binding site of M-CSF

was mapped to a similar region (Taylor *et al.*, 1994).

Structural characteristics of SCF-Kit and related ligand-receptor complexes

Kit, the receptor for SCF, is a class III receptor tyrosine kinase. This class, which includes the receptors for PDGF and M-CSF, is also closely related to the class IV receptors for FGF and the class V receptors for VEGF, Flt-3 ligand and KDR (Fantl *et al.*, 1993). The ligand-binding portions of these receptors are all composed of immunoglobulin(Ig)-like domains and the kinase domains all include kinase insert sequences. The three classes are distinguished by the number of Ig repeats (five for class III, three for class IV and seven for class V) and by the length of kinase insert, which corresponds to an excursion between two helices of the kinase structure. These Ig-like receptors share similar signal transduction pathways, chromosomal localization and gene organization (Rousset *et al.*, 1995), but their ligands come with completely unrelated topologies as typified by VEGF (cystine knot) on the one hand, versus M-CSF, SCF and Flt-3 ligand (helical cytokine) on the other. Even receptors of the same class have unrelated ligands; thus both SCF and PDGF use class III receptors and VEGF anf Flt-3 ligand use class V receptors. The amino acid sequences of the ligands are extremely dissimilar

even when the fold is the same, as for PDGF vs. VEGF (25% identity) and M-CSF vs SCF (14% identity).

Although Ig-like receptors have very similar kinase portions (70% amino acid sequence identity between III and V) and about 50% identity for III or V with IV) their Ig-like domains are dissimilar in sequence both between repeats within a molecule and also at comparable positions between different receptors. (Rousset et al., 1995) Nevertheless, there are features of the receptor-ligand interaction that the class III and class V receptors have in common. First, for every studied example, the ligand binding function has been localized to the first three Ig-like domains and, where defined, to domains D2 and D3 specifically (Heidaran *et al.*, 1990; Blechman *et al.*, 1993; Lev *et al.*, 1993; Wang *et al.*, 1993; Davis-Symyth *et al.*, 1996; Barleon et al., 1997). Secondly, the ligands for all of these receptors are functional as dimers; M-CSF, VEGF and PDGF are covalently dimers, while SCF and Flt-3 ligand are non-covalently linked dimers. In each case, signaling occurs through ligand-mediated receptor oligomerization (Heldin, 1995). For SCF-Kit, it has been shown directly by biophysical methods that complexes containing toe SCF subunits and two Kit extracellular domain molecules can form in solution (Philo et al., 1996). The genetic organization of these receptor genes has the placements and phases

of introns in common (Agnes *et al.*, 1997) and the extracellular domains can be recognized from sequence motifs as telokin-like, I-set members of the Ig superfamily (Bateman and Chothia, 1995; Harpaz and Chothia, 1994).

The structure of domain D2 of Flt-1 receptor in complex with VEGF (Wiesmann *et al.*, 1997) provides a template for ligand interactions with PDGF-related receptors. Wiesmann *et al.* (1997) modeled the interaction of VEGF with D1D2D3D4(Flt-1) and discussed the likelihood that other ligand complexes with class III and class V receptors may be similar. In lightof the structure of SCF and the identified location of receptor-binding sites, the SCF-Kit complex is modeled herein.

The D2(Flt-1) domain is similar in structure to telokin, as predicted (Harpaz and Chothia, 1994), and thereby also to both domains in the structure of vascular cell adhesion molecule (VCAM)-1 (Jones *et al.*, 1995). To test the validity of VCAM-1 as a model for D2D3(Flt-1) and D2D3 (Kit), used herein was a prediction-based threading program (Fisher and Eisenberg, 1996) to thread the sequences of the Ig-like domains of Flt-1 and Kit into the telokin and VCAM-1 structures. Fits were achieved with moderate to very high confidence of similarity. The resulting structure-based sequence alignment of D2D3(Kit) with the VCAM-1 template (five gaps) has a continuous domain boundary, and residues Cys151

and Cys183 in D2(Kit) are positioned properly to make an additional disulfide bridge between strands C and F.

Characteristics of the SCF-Kit interaction

Although it has been suggested (Matous *et al.*, 1996; Mendiaz *et al.*, 1996) that SCF may interact with its receptor in a manner analogous to the ligand-receptor interactions of another helical cytokine, growth hormone (de Vos *et al.*, 1992), an alternative mode of interaction can be contemplated given the similarities among Ig-like tyrosine-kinase receptors described above. If these similarities extend to the signaling interaction, the structure of the complex of VEGF with domain D2 of Flt-1 (Wiesmann *et al.*, 1997) should provide a template for the interaction despite the disparate structures of the ligands.

To test this hypothesis next constructed was a model of the VEGF-D2D3(Flt-1) receptor complex from a rigid-body superposition of VEGF (Muller *et al.*, 1997) and VCAM-1 such as to mimic the reported VEGF- D2(Flt-1) structure (Wiesmann *et al.*, 1997). Then, keeping the dyad-symmetric receptor pair fixed, VEGF was successively replaced with the other Ig-like receptors ligands of known three-dimensional structure: PDGF (Oefner *et al.*, 1992), M-CSF (Pandit *et al.*, 1992), and SCF (this work). Each was placed on the dyad axis and

positioned to optimize contacts between the VEGF-binding site on the receptor and the putative receptor-binding regions of the ligands. Remarkably, these disparate dimeric ligands have similar spacings between binding sites and a satisfactory fit is possible for each (Figure 6). Also constructed were simple homology models of the various receptors with changes in the backbone only to accommodate insertions and deletions. The model for SCF with D2D3(Kit) shows a striking electrostatic complementarity between a highly negative binding surface on SCF and a positive surface on Kit (**Figures 7A and 7B**). The glycosylation sites on both molecules are also compatible with unimpeded interaction.

The Kit receptor is activated by both soluble and membrane-bound forms of SCF, and signaling from the membrane-bound form appears to be have *in vivo* roles (se Lyman and Jacobsen, 1998). Moreover, as in the case of Flt-1 (Barleon et al., 1997), the D4(Kit) may be involved in inter-receptor contacts in the signaling dimer (Blechman *et al.*, 1995) [although this proposal for Kit has been questioned (Philo *et al.*, 1996; Lemmon et al., 1997)]. The model constructed herein for the SCF-Kit complex is compatible with these properties (**Figure 7A and 7B**). The C-termini of the SCF dimer are directed oppositely from those of Kit, as would be appropriate for a cell-cell contact, and the receptor units cross naturally at D4. It is

noteworthy that the ligands of other Ig-like receptors also have membrane-bound forms (M-CSF and Flt-3 ligand) or are typically complexed to the extracellular matrix (Kawasaki and Ladner, 1990; Lyman and Jacobsen, 1998).

The ligand-receptor structures that are suggested herein for the Ig-like kinase receptors are remarkable. Despite marked differences in ligand structure as typified by VEGF(cystine knot), SCF (helical cytokine) and FGF (beta trefoil), the geometrical configurations of receptor binding sites on these ligands are alike. Coupled with features in common among the receptors and in their biology, a similar mode of ligand-receptor interaction across the Ig-like subfamily of receptor tyrosine kinases seems plausible.

SECOND SERIES OF EXPERIMENTS

Sub C19

Based on the X-ray crystallographic structure of SCF, several analogs were made and their biological activities were measured and compared to that of SCF wild type.

Analogs	Biological Activity (Approximate, compared to wild type SCF)
SCF(Y26C) disulfide linker	2 to 3 fold higher
SCF(D25C)	100 fold lower
SCF(K62C)	7 fold lower

These analogs were designed based on the structure of the dimer interface of SCF, which is a non-covalent dimer. Leu22, Pro23, Lys24, Asp25, Tyr26, Lys62 and Phe63 are in the dimer surface. The side chains of Leu22, Pro23, Tyr26, and Phe63 reside in the buried center of the dimerization site and are involved in hydrophobic interactions. The hydrophilic side chains of Lys24, Asp25 and Lys62 from each monomer residue in the solvent accessible surface, and are involved in ionic interactions. By replacing Tyr26 with Cys, [SCF(Y26C)], it was anticipated that a dimer covalently linked by a disulfide bond between the C26 residue of each monomer would form because the distance between the β carbons of the two Cys26 rresidues would be less than 3Å.

Sub C19 cont'd

Analogs	Biological Activity (Approximate, compared to wild type SCF)
SCF(K78N, N81K)	3 fold lower
SCF(R117A, I118A)	10 fold lower
SCF(E92A, S95A)	no change
SCF(D124A, K127D)	no change

These analogs were designed based on the assumption that there may be two distinct receptor binding sites, per monomer, as with growth hormone. One site would be on the face between helix A and helix C, and the other site would be on the face between helix A and helix D.

References

Agnes, F., et al. (1997) Genomic organization of the extracellular coding region of the human FGFR4 and FLT4 genes: Evolution of the genes encoding receptor tyrosine kinases with immunoglobulin-like domains. *J. Mol. Evol.*, **45**, 45-49.

Anderson, D. M., et al. (1990) Molecular cloning of mast cell growth factor, a hematopoietin that is active in both membrane bound and soluble forms. *Cell*, **63**, 235-243.

Andre, C., et al. (1992) Genomic organization of the human c-kit gene: Evolution of the receptor tyrosine kinase subclass III. *Oncogene*, **7**, 685-691.

Arakawa, T., et al. (1992) Molecular weights of glycosylated and nonglycosylated forms of recombinant human stem cell factor determined by low-angle laser scattering. *Analytical Biochem.*, **203**, 52-57.

Arakawa, T., et al. (1991) Glycosylated and unglycosylated recombinant-derived human stem cell factor are dimeric and have extensive regular secondary structure. *J. Biol. Chem.*, **266**, 18942-18948.

Aronson, H.-E. G., et al. (1994) Quantification of tertiary structural conservation despite primary

sequence drift in the globin fold. *Prot. Sci.*, **3**, 1706-1711.

Barleon, B., et al. (1997) Mapping of the sites for ligand binding and receptor dimerization at the extracellular domain of the vascular endothelial growth factor receptor flt-1. *J. Biol. Chem.*, **272**, 10382-10388.

Bateman, A. and Chothia, C. (1995) Outline structures for the extracellular domains of the fibroblast growth factor receptors. *Nature Structural Biology*, **2**, 1068-1074.

Bazan, J. F. (1991) Genetic and structural homology of stem cell factor and macrophage colony-stimulating factor. *Cell*, **65**, 9-10.

Bella, J., et al. (1998) The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. *Proc. Natl. Acad. Sci. USA*, **95**, 4140-4145.

Besmer, P. (1997) Kit-ligand-stem cell factor. In Garland, J.M., Quesenberry, P.J. and Hilton, D.J. (eds), Colony-Stimulating Factors: Molecular and Cellular Biology, 2nd edn., Marcel Dekker, Inc., New York, NY, PP. 369-404.

Blechman, J. M., et al. (1993) Soluble c-kit

proteins and antireceptor monoclonal antibodies confine the binding site of the stem cell factor. *J. Biol. Chem.*, **268**, 4399-4406.

Blechman, J. M., et al. (1995) The fourth immunoglobulin domain of the stem cell factor receptor couples ligand binding to signal transduction. *Cell*, **80**, 103-113.

Broudy, V. C. (1997) Stem cell factor and hematopoiesis. *Blood*, **90**, 1345-1364.

Brünger, A. T., et al. (1987) Crystallographic *R*-factor refinement by molecular dynamics. *Science*, **235**, 458-460.

Casasnovas, J. M., et al. (1997) Crystal structure of ICAM-2 reveals a distinctive integrin recognition surface. *Nature*, **387**, 312-315.

Casasnovas, J. M., et al. (1998) A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proc. Natl. Acad. Sci. USA*, **95**, 4134-4139.

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, D, **50**, 252-270.

Davis-Symyth, T., et al. (1996) The second immunoglubulin-like domain of the VEGF tyrosine

kinase receptor Flt-1 determines ligand binding and may initiate a signal transduction cascade. *EMBO J.*, **15**, 4919-4927.

de Vos, A. M., et al. (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science*, **255**, 306-312.

Diederichs, K., et al. (1991) Novel fold and putative receptor binding site of granulocyte-macrophage colony-stimulating factor. *Science*, **254**, 1779-1782.

DiGabriele, A. D., et al. (1998) Structure of a heparin-linked biologically active dimer of fibroblast growth factor. *Nature*, **393**, 812-817.

DiMario, J., et al. (1989) Fibroblast growth factor in the extracellular matrix of dystrophic (mdx) mouse muscle. *Science*, **244**, 688-690.

Ealick, S. E., et al. (1991) Three-dimensional structure of recombinant human interferon-gamma. *Science*, **252**, 698-702.

Evans, S. V. (1993) SETOR: Hardware lighted three dimensional solid model representations of macromolecules. *J. Mol. Graphics*, **11**, 134-138.

Fantl, W. J., et al. (1993) Signalling by receptor tyrosine kinases. *Annu. Rev. Biochem.*, **62**, 453-481.

Fenstermaker et al., (1993) A cationic region of the platelet-derived growth factor (PDGF) A-chain (Arg159-Lys160nLys161) is required for receptor binding and mitogenic activity of the PDGF-AA homodimer., *J. Biol. Chem.*, **268**, 10482-10489.

Fisher, D. and Eisenberg, D. (1996) Fold recognition using sequence-derived predictions. *Prot. Sci.*, **5**, 947-955.

Fitzgerald, P. M. D. (1988) MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallogr.*, **21**, 273-278.

Flanagan, J. G. and Leder, P. (1990) The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. *Cell*, **63**, 185-194.

Fukuda, N., et al. (1997) Role of long-form PDGF A-chain in the growth of vascular smooth muscle cells from spontaneously hypertensive rats. *Am. J. Hypertens.*, **10(10 Pt 1)**, 1117-1124.

Galli, S. J., et al. (1994) The kit ligand, stem cell factor. *Adv. Immunol.*, **55**, 1-96.

Gewirth, D. (1995) *The HKL Manual*. Yale University, New Haven, Connecticut.

Glaspy, J. (1996) Clinical applications of stem

cell factor. *Curr. Opin. Hematol.*, 3, 223-229.

Harpaz, Y. and Chothia, C. (1994) Many of the immunoglobulin superfamily domains in cell adhesion molecules and surface receptors belong to a new structural set which is close to that containing variable domains. *J. Mol. Biol.*, **238**, 528-539.

Heidaran, M. A., et al. (1990) Chimeric alpha- and beta-platelet-derived growth factor (PDGF) receptors define three immunoglobulin-like domains of the alpha-PDGF receptor that determine PDGF-AA binding specificity. *J. Biol. Chem.*, **265**, 18741-18744.

Heldin, C.-H. (1995) Dimerization of cell surface receptors in signal transduction. *Cell*, **80**, 213-223.

Hendrickson, W. A. (1979) Transformations to optimize the superposition of similar structures. *Acta Cryst*, A, **35**, 158-163.

Hendrickson, W. A. et al. (1985) Direct phase determination based on anomalous scattering. *Methods Enzymology*, **115**, 41-55.

Houck, K. A., et al. (1991) The vascular endothelial growth factor family: identification of a fourth molecular species and characterization of alternative splicing of RNA. *Mol. Endocrinol.*, **5**,

1806-1814.

Hsu, Y.-R., et al. (1996) *In vitro* methionine oxidation of Escherichia coli-derived human stem cell factor: effects on molecular structure, biological activity, and dimerization. *Protein Sci.*, **5**, 1165-1173.

Hsu, Y.-R., et al. (1998) Selective deamidation of recombinant human stem cell factor during *in vitro* aging: isolation and characterization of the aspartyl and isoaspartyl homodimers and heterodimers. *Biochemistry*, **37**, 2251-2262.

Huang, E., et al. (1990) The hematopoietic growth factor KL is encoded by the *Sl* locus and is the ligand of the c-kit receptor, the product of the *W* locus, Cell, **63**, 225-233.

Jones, D. T., et al. (1992) A new approach to protein fold recognition. *Nature*, **358**, 86-89.

Jones, E. Y., et al. (1995) Crystal structure of integrin-binding fragment of vascular cell adhesion molecule-1 at 1.8 angstroms resolution. *Nature*, **373**, 539-544.

Jones, T. A. (1992) a, yaap, asap, @#*? A set of averaging programs, In S. Bailey, Hubbard, R. & Waller, D. (ed.) *Molecular Replacement, Proceedings*

of the CCP4 Study Weekend, Daresbury Laboratory, Warrington, UK, pp. 92-105.

Jones, T. A., et al. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, A, **47**, 110-119.

Kawasaki, E. S. and Ladner, M. B. (1990) Molecular biology of macrophage colony-stimulating factor. *Immunol. Ser.*, **49**, 155-176.

Kelvenbach, C. G., et al. (1991) Densimetric determination of carbohydrate content in glycoproteins. *J. Biochem. Biophys Methods*, **23**, 295-300.

Kleywegt, G. J. and Jones, T. A. (1994) Halloween ...masks and bones. In Bailey, S. Hubbard, R., and Waller, D. (eds), From First Map to Final Model, Proceedings of the CCP4 Study Weekend. Daresbury Laboratory, Warrington, UK, pp 59-66.

Koths, K. (1997) Structure-function studies on human macrophage colony-stimulating factor. *Mol. Reprod. Dev.*, **46**, 31-37.

Langley, K. E., et al. (1992) Purification and characterization of soluble forms of human and rat stem cell factor recombinantly expressed by *Escherichia Coli* and by Chinese hamster ovary

cells. *Arch. Biochem. Biophys.*, **295**, 21-28.

Langley, K. E., et al. (1994) Properties of variant forms of human stem cell factor recombinantly expressed in *Escherichia coli*. *Arch. Biochem. Biophys.*, **311**, 55-61.

Laskowski, R. A., et al. (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283-291.

Lemmon, M.A., et al. (1997) Kit receptor dimerization is driven by bivalent of stem cell factor, *J. Biol. Chem.*, 272, 6311-6317.

Lev, S., et al., (1992) Dimerization and activation of the Kit receptor by monovalent and bivalent binding of the stem cell factor. *J. Biol. Chem.*, **267**, 15970-15977.

Lev, S., et al., (1993) Interspecies molecular chimeras of Kit help define the binding site of the stem cell factor. *Mol. Cell. Biol.*, 13, 2224-2234.

Lev, S., et al., (1994) Steel factor and *c-kit* protooncogene: genetic lessons in sugnal transduction, *Crit. Rev. Oncog.*, 5, 141-168.

Lu, H. S., et al., (1991) Amino acid sequence and post-translational modification of stem cell factor isolated from buffalo rat liver cell-conditioned

medium. *J. Biol. Chem.*, **266**, 8102-8107.

Lu, H. S., et al., (1992) Post-translational processing of membrane-associated recombinant human stem cell factor expressed in Chinese hamster ovary cells. *Arch. Biochem. Biophys.*, **298**, 150-158.

Lu, H. S., et al., (1996) Isolation and characterization of a disulfide-linked human stem cell factor dimer. *J. Biol. Chem.*, **271**, 11309-11316.

Lyman, S.D, and Jacobsen, S.E.W. (1998) c-Kit ligand and Flt3 ligand: stem/progenitor cell factors with overlapping yet distinct activities, Blood, **91**, 1101-1134.

Martin, F. H., et al., (1990) Primary structure and functional expression of rat and human stem cell factor DNAs. *Cell*, **63**, 203-211.

Matous, J. V., et al., (1996) Structure-function relationships of stem cell factor: an analysis based on a series of human-murine SCF chimera and the mapping of a neutralizing monoclonal antibody. *Blood*, **88**, 437-444.

McDonald, N. Q., et al., (1995) Crystal structure of dimeric human ciliary neurotrophic factor determined by MAD phasing. *EMBO J.*, **14**, 2689-2699.

McKay, D.B. (1992) Unraveling the structure of IL-2, *Science*, **257**, 412-413.

McNiece, I. K. and Briddell, R. A. (1995) Stem cell factor. *J. Leukoc. Biol.*, **57**, 14-22.

Mendiaz, E. A., et al. (1996) Epitope mapping and immunoneutralization of recombinant human stem cell factor. *Eur. J. Biochem.*, **239**, 842-849.

Milburn, M. V., et al. (1993) A novel dimer configuration revealed by the crystal structure at 2.4 Å resolution of human interleukin-5. *Nature*, **363**, 172-176.

Muller, Y. A., et al. (1997) Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site. *Proc. Natl. Acad. Sci. USA*, **94**, 7192-7197.

Neufeld, G., et al. (1994) Vascular endothelial growth factor and its receptors. *Prog. Growth Factor Res.*, **5**, 89-97.

Nicholls, A., et al. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281-296.

Oefner, C., et al. (1992) Crystal structure of human platelet-derived growth factor BB. *EMBO J.*,

11, 3921-3926.

Otwinowski, Z. (1993) Oscillation data reduction program, In N.I.L. Sawyer, and S. Bailey (eds.) *Data Collection and Processing*, Science and Engineering Research Council, Warrington, UK, Vol. pp. 55-62.

Pähler, A., et al., (1990) A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Cryst*, A, **46**, 537-540.

Pandit, J., et al., (1992) Three-dimensional structure of dimeric huam recombinant macrophage colony-stimulating factor. *Science*, **258**, 1358-1362.

Philo, J. S., et al., (1996) Human stem cell factor dimer forms a complex with two molecules of the extracellular domain of its receptor, Kit. *J. Biol. Chem.*, **271**, 6895-6902.

Qiu, F., et al., (1988) Primary structure of c-kit: relationship with the CSF-1/PDGF receptor kinase family-oncogenic activation of v-kit involves deletion of extracellular domain and C terminus. *EMBO J.*, **7**, 1003-1011.

Raines, E.W. and Ross, R. (1992) Compartmentalization of PDGF on extracellular binding sites dependent on exon-6-encoded

sequences. *J. Cell Biol.*, **116**, 533-543.

Read, R. J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr*, **A**, **42**, 140-149.

Rousset, D., et al., (1995) Molecular evolution of the genes encoding receptor tyrosine kinase with immunoglobulinlike domains. *J. Mol. Evol.*, **41**, 421-430.

Rozwarski, D. A., et al., (1994) Structural comparisons among the short-chain helical cytokines. *Structure*, **2**, 159-173.

Russell, E.S. (1979) Hereditary anemias of the mouse: a review for geneticists. *Adv. Genet.*, **20**, 357-459.

Shrake, A. and Rupley, J. A. (1973) Environment and exposure to solvent of protein atoms: lysozyme and insulin. *J. Mol. Biol.*, **79**, 351-371.

Shull, R.M. et al., (1992) Canine stem cell factor (c-kit ligand) supports the survival of hematopoietic progenitors in long-term canine marro

Stanley, E. R. and Guilbert, L. J. (1981) Methods for the purificaiton, assay, characterization and target cell binding of a colony stimulating factor (CSF-1). *J. Immunol. Methods*, **42**, 253-284.

Sprang, S.R. and Bazan J.F. (1993) Cytokine structural taxonomy and mechanisms of receptor engagement, *Curr. Opin. Struct. Biol.*, **3**, 815-827.

Stein, J., et al. (1990) Direct stimulation of cells expressing receptors for macrophage colony-stimulating factor (CSF-1) by a plasma membrane-bound precursor of human CSF-1. *Blood*, **76**, 1308-1314.

Tan, K., et al., (1998) The structure of immunoglobulin superfamily domains 1 and 2 of MAdCAM-1 reveals novel features important for integrin recognition. *Structure*, **6**, 793-801.

Taylor, E. W., et al., (1994) Structure-function studies on recombinant human macrophage colony-stimulating factor (M-CSF). *J. Biol. Chem.*, **269**, 31171-31177.

Toksoz, D., et al., (1992) Support of human hematopoiesis in long-term bone marrow cultures by murine stromal cells selectively expressing the membrane-bound and secreted forms of the human homolog of the steel gene product, stem cell factor. *Proc. Natl. Acad. Sci. USA*, **89**, 7350-7354.

Wang, J. H., et al., (1995) The crystal structure of an N-terminal two-domain fragment of vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on the domain 1 C-D loop can inhibit

VCAM-1-alpha 4 integrin interaction. *Proc. Natl. Acad. Sci. USA*, **92**, 5714-5718.

Wang, Z., et al., (1993) Identification of the ligand-binding regions in the macrophage colony-stimulating factor receptor extracellular domain. *Mol. Cell. Biol.*, **13**, 5348-5359.

Wiesmann, C., et al., (1997) Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. *Cell*, **91**, 695-704.

Wlodawer, A., et al., (1992) Crystal structure of human recombinant interleukin-4 at 2.25 Å, *FEBS Lett.*, **309**, 59-64.

Zdanov, A., et al., (1995) Crystal structure of interleukin-10 reveals the functional dimer with an unexpected topological similarity to interferon gamma. *Structure*, **3**, 591-601.

Zimmer, Y., et al., (1993) Multiple structural elements determine lignad binding of fibroblast growth factor receptors: Evidence that both Ig domain 2 and 3 define receptor specificity. *J. Biol. Chem.*, **268**, 7899-7903.